CS Research Methods

WM9P1

Probably the most important module for successfully completing your end-year project

Dr. Elzbieta Titis, CSC WMG

How this module is structured

- 3 x 1-week block in week 2, 11 and 27.

- October Block (wc 7.10): Research process
- December Block (wc 9.12): Research methods
  - Data analysis
  - CWs and MSc Project
- March Block (wc 31.03): Research ethics
  - Ethical approval process

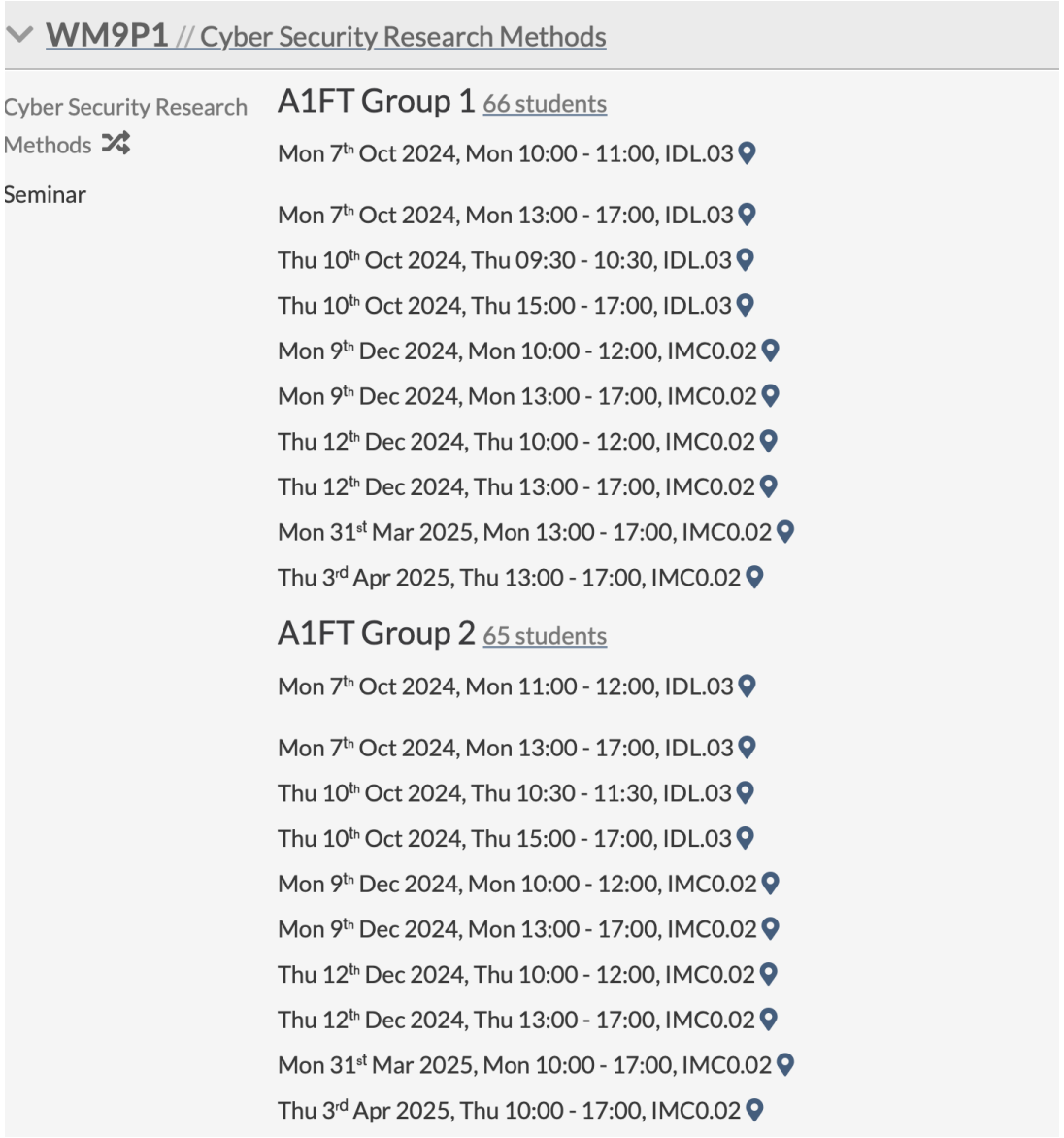
Plan for October

2x 1h class

6h seminar (groupwork)

Independent study time (work on CW1)

- Introduction to cyber security science.
  - The concept, scientific method.
- Research process:
  - Research objectives/questions, literature review (Thursday), data collection, data analysis (December).
- Self-study time:
  - Apply seminar tasks to your PMA topic (CW1). Report results in Latex and share with module tutor (Moodle) by 4th December. Prepare to discuss in December (formative feedback).



Types of research

- Applied Research
- Basic Research
- Correlational Research
- Descriptive Research
- Ethnographic Research
- Experimental Research
- Exploratory Research
- Grounded Theory
- Historical Research
- Phenomenological Research
- Quantitative research
- Qualitative research

What Is Cyber Security Science?

- The technology and practices used to protect computer networks, computers, and data from harm.
  - “How and why” we build or deploy security controls.
- Cyber security is an applied science.
  - The goal of the work is to discover how to meet a specific need.

The scientific method

- Motivations are to uncover new truths and to root out error.
  - “What if ” questions.
- Structured way of investigating the world.
  - Based on scientific ideas or methods, specifically empirical method.
- An empirical method in which the steps are based on observation, investigation, or experimentation.
  1. Formulating a question from previous observations, measurements, or experiments;
  2. Induction and formulation of hypotheses;
  3. Making predictions from the hypotheses;
  4. Experimental testing of the predictions;
  5. Analysis and modification of the hypotheses.

Principles of the scientific method

Objective	•A fair, objective experiment is free from bias and considers all the data (or a representative sample), not just data that validates your hypothesis.
Falsifiable	•It must be possible to show that your hypothesis is false.
Reproducible	•It must be possible to reproduce your results.
Predictable	•The results can be used to predict future outcomes in other situations.
Verifiable	•Nothing is accepted until verified through adequate observations or experiments.

Research process

Plan your project

1. Select the research area
2. Formulate research aim, objectives and research questions OR develop hypotheses
3. Conduct the literature review
4. Select data collection methods (collect primary data)
5. Data analysis
6. Reach conclusions
7. Complete the research & Prepare first draft

You will have to include the explanation of research process in your methodology chapter.

Research questions and objectives

- A research question is a question that a study or research project aims to answer.
- Research objectives describe what you expect to achieve by a project.
  - Once you have developed your project aim, you can start to develop objectives, and later also choose a method for each objective.

Four-step process:

- Develop aim(s) and objectives;
- Find potential methods;
- Choose methods;
- Present details of the chosen set of methods.

Good and bad questions

Example 1

- **Bad:** How does social media affect people's behaviour?
- **Good:** What effect does the daily use of YouTube have on the attention span of children aged under 16?
- Why?

Example 2

- Has there been an increase in childhood obesity in the US in the past 10 years?
- How have school intervention programs and parental education levels affected the rate of childhood obesity among 1st to 6th-grade students?
- Bad/good? Why?

SMART objectives

Specific

Measurable

Achievable

Realistic

Timely

Creating a hypothesis

- A hypothesis is a statement and suggested explanation.
  - Scientific experimentation, investigation, or observation used to show support or rejection for the hypothesis.
  - Must be testable; experiments help to decide whether hypothesis is true.

- The null hypothesis, often written as H0, is the claim that there is no relationship between two variables.
  - Offered with an alternative hypothesis called H1.
  - The null hypothesis is assumed to be true, meaning you must show evidence to prove a relationship that rejects or disproves the null hypothesis.

How to write null and alternative hypotheses

- To write your research question, null hypothesis, and alternative hypothesis, fill in the following sentences with your variables:
  - Does independent variable affect dependent variable?
    - Null hypothesis (H_0): Independent variable does not affect dependent variable.
    - Alternative hypothesis (H_A): Independent variable affects dependent variable.
  - Once you know the statistical test you will be using, you can write hypotheses in a more specific way.

Statistical test	Null hypothesis (H_0)	Alternative hypothesis (H_a)
Two-sample *t* test *or* One-way ANOVA with two groups	The mean dependent variable does not differ between group 1 (μ_1) and group 2 (μ_2) in the population; $\mu_1 = \mu_2$.	The mean dependent variable differs between group 1 (μ_1) and group 2 (μ_2) in the population; $\mu_1 \neq \mu_2$.
One-way ANOVA with three groups	The mean dependent variable does not differ between group 1 (μ_1), group 2 (μ_2), and group 3 (μ_3) in the population; $\mu_1 = \mu_2 = \mu_3$.	The mean dependent variable of group 1 (μ_1), group 2 (μ_2), and group 3 (μ_3) are not all equal in the population.
Pearson correlation	There is no correlation between independent variable and dependent variable in the population; $\rho = 0$.	There is a correlation between independent variable and dependent variable in the population; $\rho \neq 0$.
Simple linear regression	There is no relationship between independent variable and dependent variable in the population; $\beta_1 = 0$.	There is a relationship between independent variable and dependent variable in the population; $\beta_1 \neq 0$.
Two-proportions *z* test	The dependent variable expressed as a proportion does not differ between group 1 (p_1) and group 2 (p_2) in the population; $p_1 = p_2$.	The dependent variable expressed as a proportion differs between group 1 (p_1) and group 2 (p_2) in the population; $p_1 \neq p_2$.

Data collection

- A systematic process of gathering observations or measurements.
- Consider:
  - The aim of the research.
  - The type of data to collect.
  - The methods and procedures you will use to collect, store, and process the data.
    - Turn abstract conceptual ideas into measurable observations (operationalization).
    - Determine how you recruit participants or obtain measurements (sampling).

Method	When to use	How to collect data
Experiment	To test a causal relationship.	Manipulate variables and measure their effects on others.
Survey	To understand the general characteristics or opinions of a group of people.	Distribute a list of questions to a sample online, in person, or over the phone.
Interview/focus group	To gain an in-depth understanding of perceptions or opinions on a topic.	Verbally ask participants open-ended questions in individual interviews or focus group discussions.
Observation	To understand something in its natural setting.	Measure or survey a sample without trying to affect them.
Ethnography	To study the culture of a community or organisation first-hand.	Join and participate in a community and record your observations and reflections.
Archival research	To understand current or historical events, conditions, or practices.	Access manuscripts, documents, or records from libraries, depositories, or the internet.
Secondary data collection	To analyse data from populations that you can't access first-hand.	Find existing datasets that have already been collected, from sources such as government agencies or research organisations.

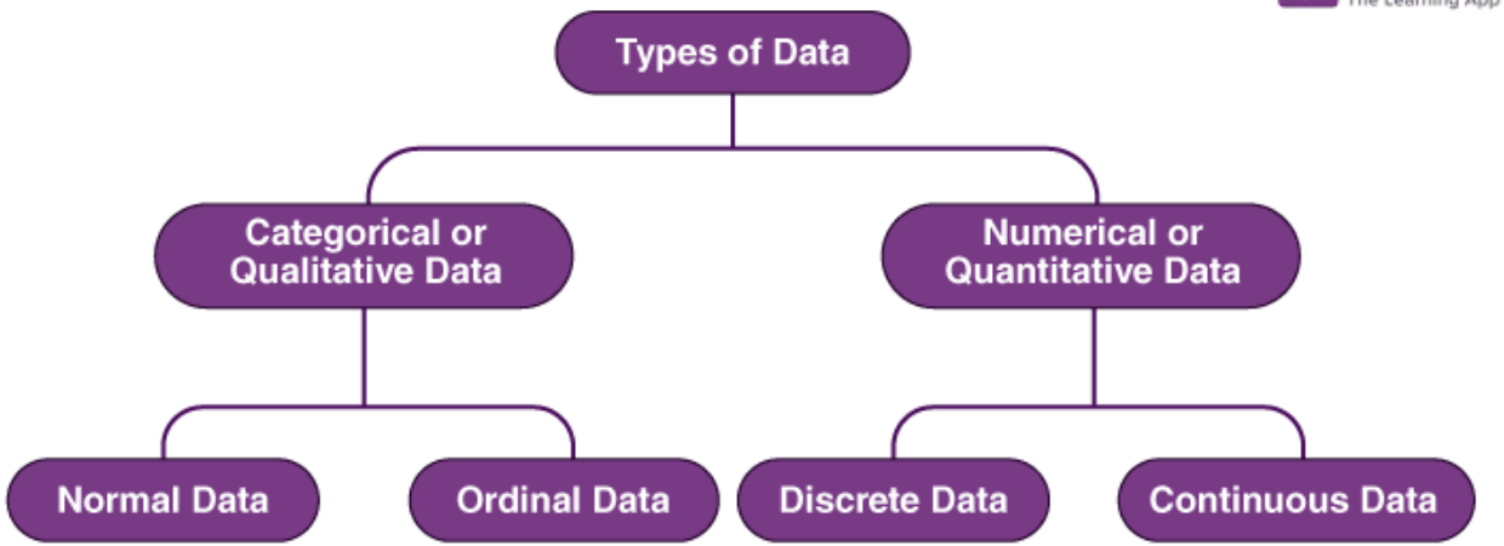
Sampling

- Allows to obtain data systematically.
- Involves defining a population, the group you want to draw conclusions about, and a sample, the group you will actually collect data from.
- Your sampling method will determine how you recruit participants or obtain measurements for your study.
  - Probability and non-probability sampling methods.
- The sample will be used to make inferences about the population.
  - Probability samples are more generalizable than non-probability samples, meaning the results can be readily applied to the greater population.
- Consider factors like the required sample size, accessibility of the sample, and time frame of the data collection.
  - There is no specific value that gives a representative sample size compared to the population. You need to be able to justify why a particular size of sample is good or bad.
  - As a rule of thumb, the bigger the sample, the more reliable it will be.

Types of data

- **Primary data** – data collected from an original source.
- **Secondary data** – data collected from a secondary source.

- **Qualitative data** – non-numerical data.
- **Quantitative data** – numerical data.



Quantitative vs Qualitative (categorical)

Types of data

Discrete vs continuous variables

Type of variable	What does the data represent?
Discrete variables (aka integer variables)	Counts of individual items or values.
Continuous variables (aka ratio variables)	Measurements of continuous or non-finite values.

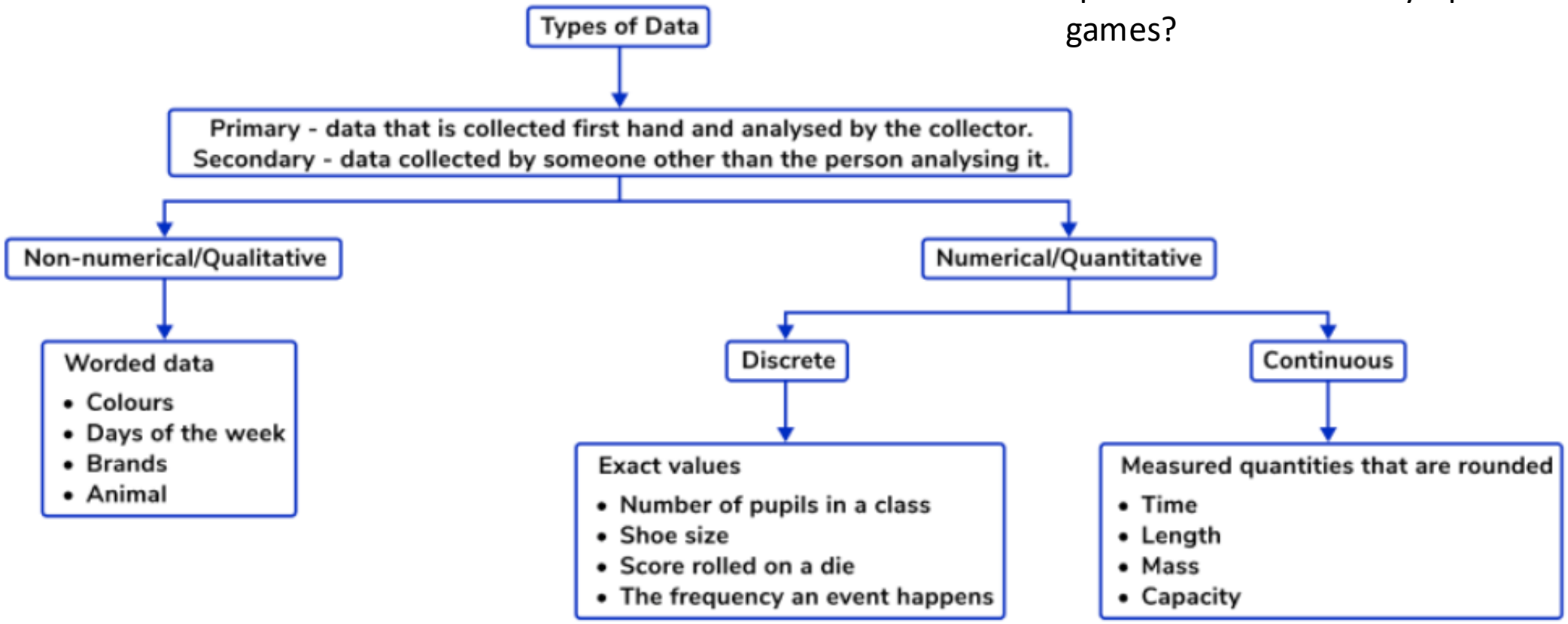
Quantitative

Binary vs nominal vs ordinal variables

Type of variable	What does the data represent?
Binary variables (aka dichotomous variables)	Yes/no outcomes.
Nominal variables	Groups with no rank or order between them.
Ordinal variables	Groups that are ranked in a specific order.

Categorical

How to recognise **types of data**:



- Survey on favourite fizzy drinks?
- Winning times of the 100-metre sprint for the last 10 Olympic games?

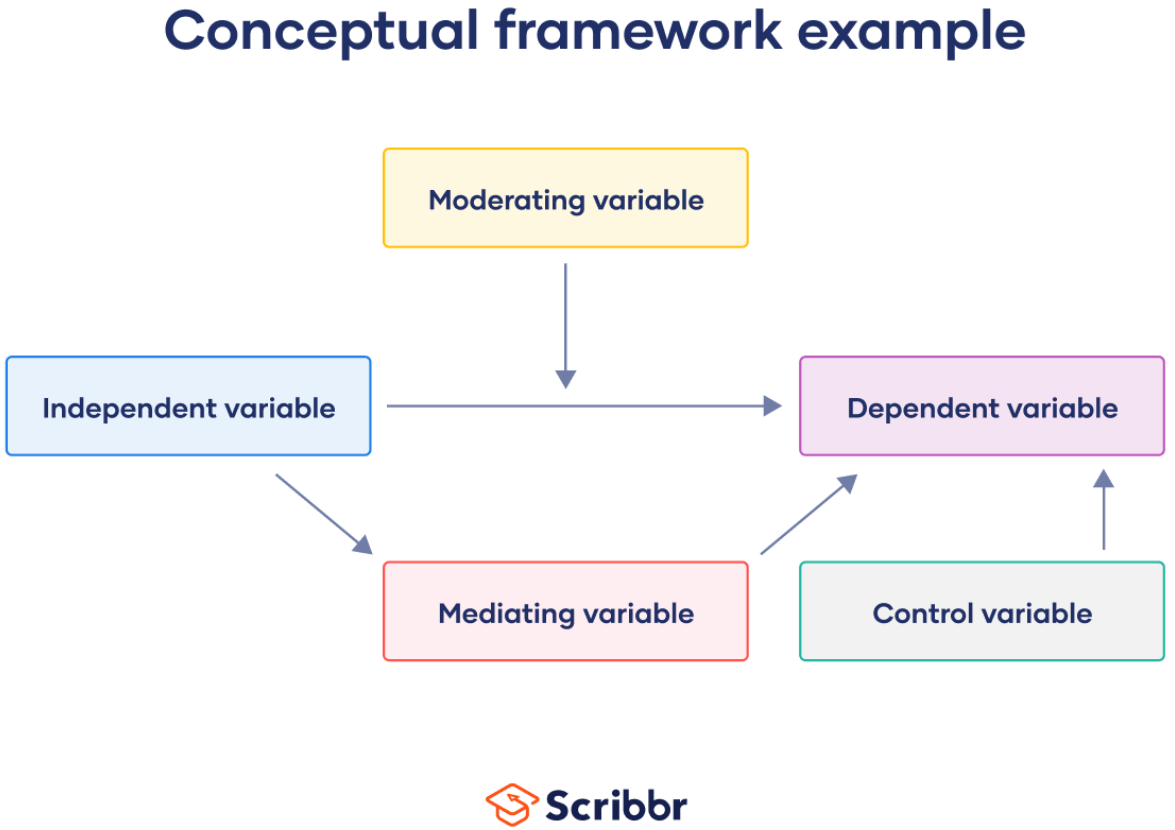
Independent vs dependent variables

Type of variable	Definition
Independent variables (aka treatment variables)	Variables you manipulate in order to affect the outcome of an experiment.
Dependent variables (aka response variables)	Variables that represent the outcome of the experiment.
Control variables	Variables that are held constant throughout the experiment.

Type of variable	Definition
Confounding variables	A variable that hides the true effect of another variable in your experiment. This can happen when another variable is closely related to a variable you are interested in, but you haven't controlled it in your experiment.
Latent variables	A variable that can't be directly measured, but that you represent via a proxy.
Composite variables	A variable that is made by combining multiple variables in an experiment. These variables are created when you analyse data, not when you measure it.

Why do I need to know types of data/variables

- Based on the data you want to collect you can decide on suitable method(s).
- Once you have defined your independent and dependent variables and determined whether they are categorical or quantitative, you will be able to choose the correct statistical test.
- It is also useful to map relationships between your variables to define the relevant objectives for your research process and maps out how they come together to draw coherent conclusions….
    - This is known as conceptual framework.



Example dataset exercise

You want to test whether some plant species are more salt-tolerant than others. Some key variables you might measure include the amount of salt you add to the water, the species of plants being studied, and variables related to plant health like growth and wilting.

- What is/are independent/dependent variable(s)?
- Identify the following types of variables: Nominal, continuous, ordinal and binary variables.

Sample	Plant species	Salt added	Starting height in cm	Growth in cm	Wilting (rank 0-10)	Survival (0=Survived; 1=Died)
1	A	0	12	…	…	0
2	A	100	13	…		0
3	A	200	11			1
4	B	0	25			…
5	B	100	26			
6	B	200	25			