Using the Titanic datasets to teach mixed methods data analysis
(WORKING PAPER – unpublished document)

Jörg Stolz[1], Anaïd Lindemann[1]

1. Introduction

Mixed methods research has made much progress in recent decades, but there is still a lack of highly visible examples and publicly available datasets that may be used for learning and teaching mixed methods (Bazeley 2003; Bryman 2008; Creswell et al. 2003). There are hundreds of free mono-quantitative and mono-qualitative datasets on the Internet or in specific data-repositories, but very few mixed methods datasets. We do not claim that we can solve the situation here, but wish merely to make a start. We add one mixed methods dataset that researchers may download and use freely for their learning and teaching. While this is only one example, we think it is a good one for three reasons.

First, students are fascinated both by the historical case of the *Titanic* and by the spectacular correlations between gender/class and likelihood of survival. Second, the example shows that a mixed methods analysis is actually superior to a mono-method analysis (Stolz and Lindemann 2019; Stolz, Lindemann and Antonietti 2018), and is therefore very useful to show the specific added value of mixed methods. Third, this is all the more noteworthy because the quantitative Titanic data are often used for the mono-method teaching of statistics in all major packages (SAS, R, SPSS, STATA) (Bellocco and Algeri 2013; Kohler and Kreuter 2017; Landau and Everitt 2004).

This paper has two goals: (1) to present the linked Titanic datasets; and (2) to present a three-hour exercise with the Titanic datasets that can be used to learn and teach mixed methods.

Since we have been using the Titanic example in our teaching both for beginners and advanced students for years, we have experience of the kinds of questions that students might have and the kinds of insights that are possible. We have therefore tried to integrate this knowledge, and have been inspired in doing so somewhat by Strauss (2003 (1987)).

As will become clear, the exercises reflect to a certain extent our views of the mixed methods approach and its philosophical background – for example, that mixed methods may use a realist philosophical background, and that one central rationale of mixed methods can be to give more valid answers to research questions by eliminating validity threats from other methods (Kelle 2001; Kelle 2007; Maxwell and Mittapalli 2010; Stolz 2016; Stolz 2017; Stolz et al. 2016; Stolz and Lindemann 2019; Stolz, Lindemann and Antonietti 2018). However, we believe that the example of the Titanic may also be useful for mixed methods researchers with other philosophical, theoretical, or methodological leanings. These researchers may want to adapt several elements of the exercises below.

---

[1] University of Lausanne, Lausanne, Switzerland

Corresponding author:
Jörg Stolz, ISSR, University of Lausanne, 1015 Lausanne, Switzerland
Email: Joerg.stolz@unil.ch

2. The Titanic Datasets

The Titanic datasets consist of a quantitative dataset (n = 2207) and a qualitative dataset of testimonies provided by the survivors (N = 214). These two datasets are linked perfectly by a variable indicating the names of the survivors.

2.1 The quantitative dataset

The initial version of the quantitative dataset was created by Frey et al. (2011) from the Encyclopedia Titanica.[i] We cross-checked and enhanced this dataset with a number of variables in light of its use in a mixed methods study. More specifically, we added variables on the time when individuals boarded a lifeboat[ii], the side of the boat from where individuals boarded the lifeboat, the order in which the lifeboats left the *Titanic*, and whether or not the individuals gave a testimony.

The quantitative dataset contains the following variables (for all details, see the Appendix):
- Dependent variables: Survived/perished, Time of boarding a lifeboat.
- Independent variables: Age, Sex, Class/Crew, Country of origin, Social ties.
- Additionally, there are some context variables: Side of the boat from which individuals boarded a lifeboat (port/starboard), Boat number, Order in which the lifeboats left the *Titanic*, name of the individual, ID, as well as a dummy variable indicating whether or not individuals gave a testimony after the tragedy.

The quantitative dataset is available as an SPSS file or a CSV file. Readers who use R can use the R syntax provided to label the variables of the CSV file.

2.1 The qualitative dataset

The qualitative dataset contains the testimonies of 214 survivors. These testimonies have been taken from the Encyclopedia Titanica[iii] and the British and American trial proceedings.[iv] The survivor testimonies were grouped according to the lifeboat that enabled the individual to survive. These testimonies come in a variety of forms: a. interviews given to journalists; b. testimonies given at the trial proceedings; c. accounts provided by journalists of what survivors had told them; d. letters to family and friends; e. affidavits. The testimonies were given at very different times, in different contexts, to different publics, and are of very different lengths (the shortest contains a few sentences, while the longest contains more than 280 pages).[v]

The qualitative dataset comes in two forms:
(1) In a number of text files, ordered according to the lifeboats in which the survivors were rescued.
(2) As a coded MAXQDA file. In Textbox 1, we present a selection of the code list (for the full code list as made public, see the Appendix). On level 1, we distinguish "Filling rules", "Authority acceptance", "Way to the boat deck", "Attribute nationality", and "Social network". Every code has a specific coding rule that can be inspected by looking at the code memo in MAXQDA (see also the Appendix). Many codes (such as filling rules and authority acceptance) come in two forms: whether the respondent *experienced* the attribute of the situation when entering the lifeboat herself, or whether she *observed* this attribute of a situation in another context. This allows us to quantify the attributes experienced and cross-tabulate them with the variables of boat side, lifeboat, gender, and class. To create the coding scheme, we used "game heuristics", an inductive method that starts out with the idea that social phenomena are interlinked "social games", and that codes have the goal to reveal "game

mechanisms" that may explain a given social outcome. The specifics of this method are described in Stolz and Lindemann (2019).

Textbox 1: Coding scheme (selection)

Level 1	Level 2	Level 3
Filling rules	Women & children first If no more women - fill randomly Couples first Fill with anybody Shortage of women Women hesitate/refuse to board	
Authority acceptance	Officers in charge People calm, follow orders People panic Men try to sneak into boats Men try to rush boats Officers re-establish order Officers shoot Fill from A deck	
Way to boat deck	Crowd size	Crowd_small Crowd_large
	Arrival Time on Deck	Arrives before 0.40 Arrives 0.40 - 1.20 Arrives 1.20 - 2.05 Arrives after 2.05 or not (all boats gone)
	Way to boat deck	On first decks when impact occurs Informed through crew On other decks first Difficult way to boat deck Crowd obstructs passage Crew obstructs / unhelpful Reluctant to go to boat deck Finds gates locked Finds gates open
	Age & strength	
	Attribute_nationality	
	Social_Network	Informed through social network Goes to deck with others Get on - encouraged by social network Social Network generic

---

2.3 Reliability and validity issues

Most variables in the quantitative dataset have a high level of reliability. We have a very thorough knowledge of who was on the *Titanic*, their age, sex and nationality, and the type of ticket (class) they had or crew member they were. For almost all survivors, we are quite confident about which lifeboat they boarded and at what approximate time. While the information in the quantitative dataset is of a high quality, we have to acknowledge that a lot of what is interesting remains unmeasured. For example, we do not have useful variables that give us information about the location of individuals' cabins, the time when they were informed about the emergency, or their activities and strategies thereafter.

The qualitative dataset gives us a wealth of information about many of the points that are missing in the quantitative dataset. However, this qualitative information is biased in various ways.

(1) *Selection bias*: One obvious bias is the fact that we only have the testimonies of surviving individuals, who are not a random subset of individuals (e.g. preference was given to women and those from higher social classes). Furthermore, only a subset of survivors provided a testimony, and the selection is again not random, with precedence being given to men, individuals of a high social status, and people of an Anglo-Saxon heritage. The two-stage selection process can be inspected in Table 1. This is a form of "sampling on the dependent variable", where the selection has been carried out by the natural and social process itself. It is well-known in quantitative research that "sampling on the dependent variable" is problematical, since it may lead in its extreme form to a complete lack of variance in the dependent variable or, in its mild form, to biased (underestimated) regression coefficients (King, Keohane & Verba 1994: 129).

Table 1   Passengers & crew, survivors, and those providing a testimony

	Those providing a testimony		Survivors		Passengers & crew	
	N	%	N	%	N	%
Women	93	43.2	351	49.9	485	22.0
Men	121	56.8	353	50.1	1722	78.0
1. Class	69	32.1	200	28.4	324	14.7
2. Class	34	15.8	115	16.3	285	12.9
3. Class	36	16.7	178	25.3	708	32.1
Restaurant Crew	2	0.9	3	0.4	69	3.1
Deck Crew	28	13.0	42	6.0	66	3.0
Engine Crew	17	7.9	72	10.2	325	14.7
Victualling Crew	29	13.5	94	13.4	430	19.5
English	122	57.0	293	41.6	1164	52.7
American	58	27.1	207	29.4	424	19.2
Other	34	15.9	204	29.0	619	28.0
N	214	100	704	100	2207	100

(2) *Psychological biases*: Other biases may result from the fact that individuals (a) may have had a poor recollection of the events, and all the more so since these events were clearly extremely traumatic; (b) may have confused their memories with stories that they heard later

or with films that they watched later; (c) may have adapted their stories to their audience in order to create specific effects – for example, to embellish their role or render certain points more dramatic; (d) may have withheld or invented important information, especially if they feared negative judgment on their behaviour.

While there are undoubtedly various ways in which bias can be introduced, we can also do much to counter bias; indeed, this is precisely the advantage of using mixed methods:
- Since our qualitative dataset is nested in the quantitative dataset, we can investigate the extent of selection bias concerning gender, class/crew, and nationality in Table 1.
- As survivor testimonies are qualitative accounts that all refer to the same event, they contain valuable information about what happened not only to those testifying, but also to others (both surviving and non-surviving), and about the evolving context as a whole. By triangulating different testimonies, we can often make very precise assertions about what actually happened on the boat deck, and build quite a good (albeit less reliable) picture of the ways to the boat deck that different classes took.
- Again, because both the qualitative and quantitative datasets are concerned with the same social process, we can also triangulate and cross-check results between data types.

3. A three-hour exercise to show the general usefulness of mixed methods

In what follows, we present an exercise that uses the Titanic datasets and that shows students the general usefulness of mixed methods. We often do the exercise in three hours, but the time obviously varies greatly according to how much time is taken to go into depth in specific steps.

3.1 Goal of the exercise
At the end of the exercise, students should
(1) Understand the usefulness of mixed methods, i.e. the fact that mixed methods may lead to more valid conclusions about a research question than a mono-method analysis.
(2) Understand that all hypotheses made about social mechanisms and the meaning/function of x and y variables rest on assumptions that may be wrong and that can be scrutinized with qualitative analysis.
(3) Understand that all the results of qualitative analysis – for example, typologies and perceived causal mechanisms – can be quantified. Sometimes, it is possible to generalize to a larger population, and relationships may be tested for statistical significance. Conversely, all variables in a quantitative dataset can be investigated as to their meaning and function in a "social game" and context with qualitative means.
(4) Understand how it is possible, in mixed methods, to switch iteratively between quantitative and qualitative analysis.

3.2 Prerequisites
For these exercises, students need a basic understanding of quantitative and qualitative methods, some knowledge of the statistical softwares SPSS and/or R as well as of the mixed methods software MAXQDA, and laptops that have either SPSS or R and MAXQDA installed. One of the advantages of R is that it is free. Unfortunately, there is not yet available a coded version of the Titanic dataset in a free qualitative software (like RQDA).

3.2 Introducing the exercise

(1) Have students whistle Céline Dion's "My heart will go on", or show them a few seconds of the trailer for the *Titanic* film. In this introduction part, you can also provide some factual and historical information regarding the Titanic. For example, inform them that 2207 individuals were on board at the time of the collision and that only 710 survived the shipwreck, despite the fact that the lifeboats could have saved 1178. In other words, the boats have been used at only 60% of their capacity.

(2) State the *research question*: What were the causes / mechanisms that led individuals on the *Titanic* to survive or perish? This step is important because we want to show students that mixed methods are not interesting in themselves, but only have legitimacy if they are able to give a more valid answer to a research problem than a simpler mono-method approach.

(3) Preliminary *theorizing*: have students find explanatory variables that may lead to higher or lower probability of surviving on the *Titanic*. Students come up with variables like class, gender, physical strength, social ties, location of the cabins, etc. We normally make a sketch on the blackboard of these variables. For every independent variable mentioned, we ask students to specify the "causal story" or "causal mechanism" of how exactly this explanatory variable might have influenced the response variable. For example, if they mention "class", we ask: how did that work exactly? They may say: people in first-class accommodation had more money than people in the lower-class cabins, and they may have bribed the crew to allow them onto the lifeboats. We encourage students to be as precise as possible for every assumed mechanism.

3.3 Exploratory quantitative data analysis

Have students do some *exploratory quantitative data analysis* with the variables that they have found in the previous step: class, gender, age, and survived/perished.

They may use the following syntax:

```
********************************
Textbox 2: Exploratory quantitative analysis - SPSS
********************************

*** (1) Analysis of bias in testimonies: example gender

crosstabs testimony by sex by lived
/cells count row
/statistics.

*** (2) Frequencies response variable

Frequencies lived /bar chart.

*** (3) Frequencies explanatory variables

Frequencies  sex classcrew  age_cat  boatnumber boatside child country_5cat testimony / barchart.

Graph/
histogram age.
```

```
####################################
Textbox 3: Exploratory data analysis - R
####################################

library(dplyr)
library(tidyr)
library(ggplot2)
library(forcats)
library(purrr)
```

 # read in data (your own location file)
```
Titanic_Mixed                                                                    <-
read.csv("~/Dropbox/2019_neu/02_Projects/P_Titanic/Titanic_Website/DATASET_QUAN/Titanic_Mixed.csv
")
```

 # make it a tibble
```
data <- tbl_df (Titanic_Mixed)
```

 # recode some variables

```
data <- data %>%
  mutate (lived = recode(factor(lived), "0" = "perished", "1" = "survived")) %>%
  mutate (sex = recode(factor(sex),"0" = "male", "1" = "female" )) %>%
  mutate (testimony = recode(factor(testimony), "0" = "no", "1" = "yes")) %>%
  mutate (group = recode (factor(group), "0" = "Single", "1" = "Single w/servant",
                "2" = "Couple", "3" = "Couple w/kids", "4" = "Couple w/servant",
                "5" = "Single parent w/kids", "6" = "Family w/servant", "7" = "Family/friends",
                "8" = "Crew groups", "9" = "Family/friends w/kids"))%>%
  mutate (country_5cat = recode (factor(country_5cat), "0" = "England", "1" = "Ireland", "2" = "Sweden", "3" =
"USA",
                "4" = "Others"))%>%
  mutate (country_3cat = recode (factor(country_3cat), "0" = "England", "1" = "USA", "2" = "Others"))%>%
  mutate (classcrew = recode (factor(classcrew), "1" = "1st class passenger", "2" = "2nd class passenger",
                "3" = "3rd class passenger", "4" = "A la carte crew", "5" = "Deck crew",
                "6" = "Engine crew", "7" = "Victualling crew")) %>%
  mutate (classcrew1 = recode (factor(classcrew1), "1" = "1st class passenger", "2" = "2nd class passenger",
                 "3" = "3rd class passenger", "4" = "Crew")) %>%
  mutate (age_cat = recode (factor(age_cat), "1" = "0-14", "2" = "15-30",
                "3" = "31-40", "4" = "41-50", "5" = "51-60",
                "6" = "61+", "99" = "NA"))%>%
  mutate(age = as.numeric(age))%>%
  mutate(boatorder = as.numeric(boatorder))%>%
```

```
    mutate(boatside = recode(factor(boatside), "0" = "Starboard", "1" = "Port"))

glimpse(data)
```

(1) Analysis of bias in testimonies : example gender

```
options (digits = 3)

all <- data %>%
  group_by(sex) %>%
  summarize (All = n()) %>%
  mutate (percent = All/sum(All) * 100)%>%
  select (- sex)

 survivors <- data %>%
  filter(lived == "survived") %>%
  group_by(sex) %>%
  summarize (Survivors = n()) %>%
  mutate (percent = Survivors/sum(Survivors)* 100)%>%
  select (- sex)

 testifiers <- data %>%
  filter(lived == "survived" & testimony == "yes") %>%
  group_by(sex) %>%
  summarize (Testifiers = n()) %>%
  mutate (percent = Testifiers/sum(Testifiers)*100)

bias_table <- cbind(testifiers, survivors, all)
bias_table
```

#(2) Frequencies response variable

```
data %>%
  group_by(lived)%>%
  count()%>%
  ungroup()%>%
  mutate(percent = n/sum(n))
```

(3) Frequencies explanatory variable

```
crosstabs <- function(df, Var1, Var2 ){
 df %>%
   group_by(!! Var1, !! Var2) %>%
   summarize (n = n()) %>%
   mutate (perc = n/sum(n)* 100) %>%
   print()
}

explanatory_vars <- list("sex", "classcrew", "age_cat", "boatnumber",
                "boatside", "child", "country_5cat", "testimony")

explanatory_vars1 <- list("sex")

for (i in 1: length(explanatory_vars)){
 crosstabs(data, quo(lived), quo(eval(parse(text=paste0(explanatory_vars[[i]])))))
}

glimpse(data)
```

(4) Barchart

```
data %>%
  group_by(sex)%>%
  summarize (survival_rate = mean(lived == "survived"))%>%
ggplot(aes(x = sex, y =survival_rate))+
  geom_bar(stat = "identity", fill = "darkblue")+
  labs(x = "", y = "Percentage survived", title = "Survival rate on the Titanic according to Gender")

# (4) Barchart

data$sex <- relevel(data$sex, ref = "female")

data %>%
  group_by(classcrew, sex,lived) %>%
  summarise(n = n()) %>%
  mutate (perc = n/sum(n)* 100) %>%
  filter (lived == "survived") %>%
  ggplot(aes(classcrew, perc, fill = sex))+
  geom_bar(stat = "identity", position = position_dodge(preserve = "single"))+
  labs(x = "Class & type of crew", y = "Percent",
      title = "Survival ratio according to class/crew & Sex")+
  theme(legend.title = element_blank())+
  scale_fill_manual(values=c("firebrick", "dodgerblue3"))+
  geom_text(aes(label = scales::percent(perc/100), y = perc + 2.2),
        position = position_dodge(width = 1), size = 3)
```

Have students summarize what they have learned from this exploratory analysis. They/we normally come to the following intermediate conclusions:

(1) *Analysis of bias in testimonies: example gender*. There is an important bias in our sampling of those providing a testimony – because it was only those that survived that gave a testimony! This is "selecting on the dependent variable", with all its known problems. However, there are even more instances of bias, because men and women did not survive with equal probability. The gender ratio of all individuals on the *Titanic* was 78% men to 22% women; of survivors, it was 50.1% men to 49.9% women; of those testifying, 57.5% men to 42.5% women. This means that we have a double selection process: women are much more likely to be in the survivor category, but, of the survivors, men are more likely to be in the testifier category. We urge students to keep this very bias in mind when interpreting our insights from the qualitative sample.

(2) *Response variable*. Of 2207 individuals in our dataset, 1497 (67.8%) perished, and only 710 (32.2%) survived.

(3) *Frequencies explanatory variables*. We look at the different frequency tables. We note, for example, that there were many more men (78%) on the *Titanic* than women (22%), and many more third-class (32.1%) than first-class (14.7%) and second-class (12.9%) passengers. Roughly 40% of the individuals on the *Titanic* were crew members.

(4) *Some simple crosstabs*. We look at survival by gender, class/crew, and gender*class/crew. As described in Stolz, Lindemann and Antonietti (2019: 1627), we find that (see also Figure 1):

- Women across all classes and types of crew generally have a higher likelihood of survival than men (e.g. first-class female passengers, 96.5% *vs*. first-class male passengers, 34.4%), the one exception being that male deck crew have a higher likelihood of survival than third-class female passengers.
- Higher-class passengers generally survive more often than lower-class passengers (e.g. first-class female passengers, 96.5% *vs*. second-class female passengers, 84.9%, *vs*. third-class female passengers, 48.6%), the one exception being that there is no significant difference between second-class and third-class male passengers.
- We find interactions between gender and class. There is a significant difference between first-class and second-class female passengers who survive (more than 80%) and third-class female passengers who survive (only 48.6%); for men, the major difference is between first-class male passengers who survive (34.4%) and second-class and third-class male passengers, who have very similar survival rates of 14% and 15% respectively. The men belonging to the restaurant crew (*A la carte*) have the lowest survival rate of all groups of men, with only 1.5% (both women belonging to the restaurant crew survive). Male deck crew have the highest likelihood of survival of all groups of men (63.6%).

Figure 1 : Survival ratio according to class/crew & sex
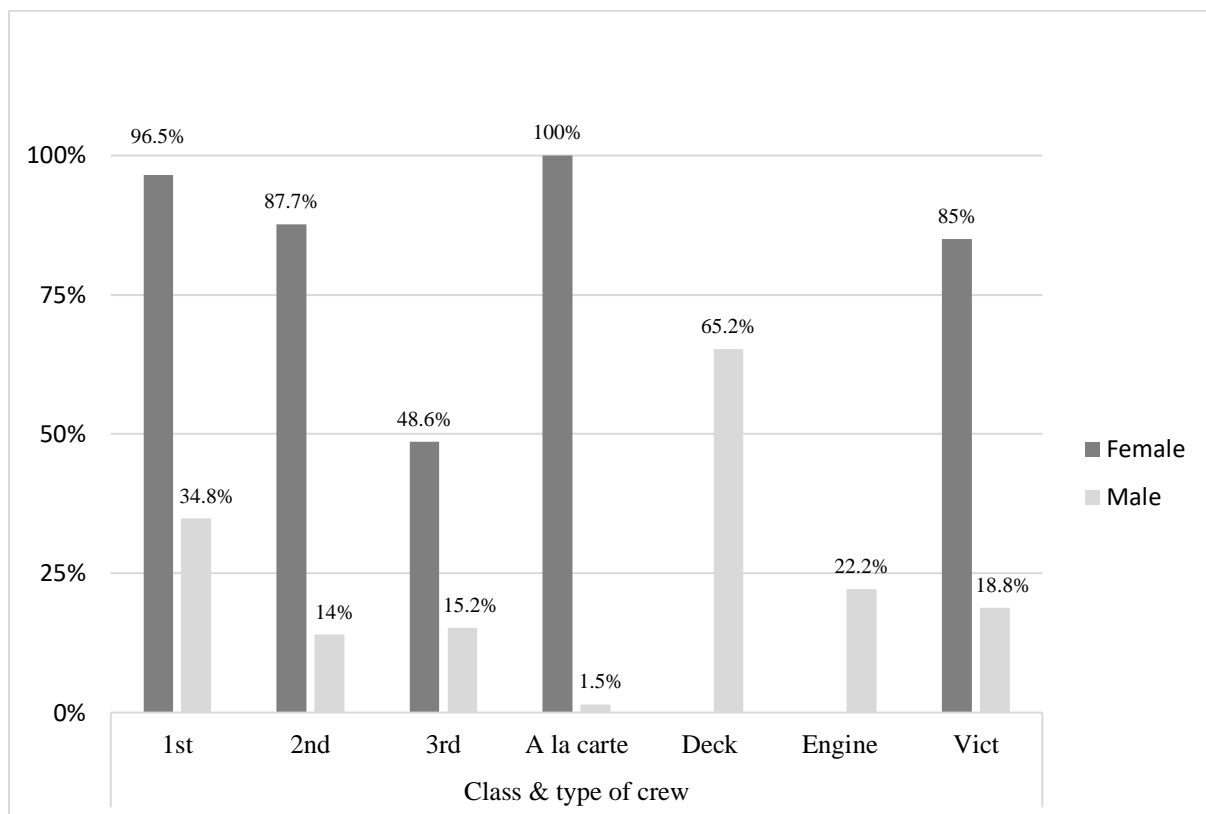
We discuss with students that these exploratory analysis have taught us much, and that they confirm or reject some of our initial hypotheses, but that much remains in the dark. Through what exact intentional and rule-governed actions and through what game mechanisms have these correlations been produced? Are there elements of the "Titanic game" that are important but not captured by the quantitative variables? To have more insight concerning these questions, we now turn to the qualitative data.

3.4 Exploratory qualitative data analysis

Have students do some exploratory qualitative analysis with MAXQDA. We introduce this part by telling students two things:

(1) They cannot possibly provide a qualitative analysis of this material in the short time available in the exercise. Good qualitative analysis means becoming thoroughly acquainted with the material as a whole, reading through all the testimonies, and carefully coding, comparing, and recoding them, etc. Here, the goal is to look at some selected testimonies and coded material to understand the way that mixed methods analysis can be put into practice.

(2) This analysis is not about how to code, since the coding is already provided. Also note that we not only coded the content of the testimonies, but also created document variables for each testimony such as "gender", "age" or "boat" of the testifier. For a description of how the coding was done, see Stolz & Lindemann (2019).

Textbox 4: Exploratory analysis in MAXQDA

(1) Compare first-class and third-class women
In MAXQDA, in the window with the list of codes, select the "Exercise" code. Right-click on it. A menu pops up. Select "Activate".
Now go to tab mixed methods. Click activation by document variables. Activate first-class women, by putting [Gender] = F AND [Class] = 1 into the right-hand box. Click "Activate". In the "View" menu, select "Selected Codings". You should now see parts of the testimonies of three first-class women. Read through these testimonies.
Now change to see three third-class women. Click activation by document variables. Activate third-class women, by putting [Gender] = F AND [Class] = 3 into the right-hand box. Click OK. You should now see parts of the testimonies of three third-class women. Read through these testimonies. What is different in the accounts provided by first- and third-class women?

(2) Compare first-class and third-class women: codings about filling rule "women and children first"
Select the Exercise code. Right-click on the "Exercise" code. Select "Deactivate". In the code "Filling rules" (experiencd), select the subcode "Women & Children first". Right-click on it. Select "Activate".
Now go to mixed methods tab. Select all first-class women as before. In the view menu, select "Selected Codings". You should now see all coded testimonies where surviving first-class women say that the filling rule "women and children first" was used for their lifeboat. Do the same thing for third-class women, and compare.

(3) Make crosstabulation Authority acceptance * Boatorder
In MAXQDA, in the document window, activate all the documents (Boat 1 to Collapsible A).
In the code window, activate all codes under "Authority acceptance".
Go to tab "Mixed Methods". Go to Crosstabulation. Click on the Variable "Boat" in the left-hand window. Click below on "Insert all values into the table". Click on the arrow. This puts all boats into the right-hand window. Now change the order of the boat (manually change the "Values" for each one and remove "boat 99") until you have the following order (order in which lifeboats left the Titanic):
[Boat] = 7
[Boat] = 5

[Boat] = 3
[Boat] = 8
[Boat] = 1
[Boat] = 6
[Boat] = 16
[Boat] = 14
[Boat] = 12
[Boat] = 9
[Boat] = 11
[Boat] = 13
[Boat] = 15
[Boat] = 2
[Boat] = 10
[Boat] = 4
[Boat] = C
[Boat] = D
[Boat] = B
[Boat] = A

Click on OK. What generalizations can be made about authority acceptance during the filling of the lifeboats?

(4) Make crosstabulation Filling rules * Boatorder

In MAXQDA, in the document window, activate all the documents (Boat 1 to Collapsible A). These documents should now appear in red.

In the code window, activate all codes under Filling rules/Experienced. These codes should now appear in red.

Go to tab "Mixed Methods". Go to Crosstabulation. Click on the Variable "Boat" in the left-hand window. Click below on "Insert all values into the table". Click on the arrow. This puts all boats into the right-hand window. Now change the order of the boat (manually change the "Values" for each one and remove "boat 99") until you have the following order (order in which lifeboats left the Titanic):

[Boat] = 7
[Boat] = 5
[Boat] = 3
[Boat] = 8
[Boat] = 1
[Boat] = 6
[Boat] = 16
[Boat] = 14
[Boat] = 12
[Boat] = 9
[Boat] = 11
[Boat] = 13
[Boat] = 15
[Boat] = 2
[Boat] = 10
[Boat] = 4
[Boat] = C
[Boat] = D
[Boat] = B
[Boat] = A

Click on OK. Did the application of filling rules change during the process of filling?

(5) Make crosstabulation Filling rules * Boatside

In MAXQDA, in the document window, activate all the documents (Boat 1 to Collapsible A). These documents should now appear in red.

In the code window, activate all codes under Filling rules/Experienced. These codes should now appear in red.

Go to tab "Mixed Methods". Go to Crosstabulation. Click on the Variable "Boatside" in the left-hand window. Click below on "Insert all values into the table". Click on the arrow. This puts all levels of "boatside" into the right-hand window. Now remove [Boatside] = 99

Click on OK. Use row percentages. Did the application of filling rules differ on port and starboard?

(6) Make crosstabulation Way to the boat deck * Classcrew

In MAXQDA, in the document window, activate all the documents (Boat 1 to Collapsible A), except "Other witnesses". These documents should now appear in red.

In the code window, activate all codes under "Arrival time on deck" and "Way to boat deck – Experienced". These codes should now appear in red.

Go to tab "Mixed Methods". Go to Crosstabulation. Click on "Class" in the left-hand window. Click below on "Insert all values into the table". Click on the arrow. This puts all classes into the right-hand window. Now change the order of the classes until you have the following order:

[Class] = 1
[Class] = 2
[Class] = 3
[Class] = Deck crew
[Class] = Restaurant crew
[Class] = Engine crew
[Class] = Victualling crew

Click on OK. Did different classes differ in terms of their experience when trying to get to the boat deck?

Have students summarize what they have learned from this exploratory qualitative analysis. They/we normally come to the following intermediate conclusions:

(1)/(2) *Compare first-class and third-class women:* Time is important. It seems that first-class passengers arrived on the boat deck (where the lifeboats were) earlier than second-class passengers, who in turn arrived earlier than third-class passengers. We point out to students that we could test this idea with our quantitative dataset in a next step. We also encourage students to dig further into the data to find the exact reasons for why this was the case.

(3) *Crosstabulation Authority acceptance * Boatorder:* The crew was important when filling the boats. This is a very important point because it means that the rational behaviour of individuals is of very limited use when explaining survival probabilities on the *Titanic*. The crew was very much in control of what happened, and the final outcome depended largely on how they put their rescuing and filling rules into practice.

(4) *Crosstabulation Filling rules * Boatorder:* The rule "Women and children first" was used in all boats, except the two last ones. The rule "If no more women - fill up with men" was used in only some boats and not others. The rule "Couples first" is only mentioned for the two first boats. The rule "Fill with anybody" is mentioned only for the two first and the two last boats. We explain to students that, with more time, one would now go deeper into the material to understand the reason for these differences. Clearly, at the beginning of the filling process, the practice was not yet clearly defined. At the end of the process, the water was already very high and an orderly filling was not possible anymore.

(5) *Crosstabulation Filling rules * Boatside:* Filling rules were different on port and starboard. Clearly, the rule "Women and children first" was interpreted differently on Port and Starboard. On Port, it was understood as "Women and children only". On Starboard it was interpreted as: "Fill up with women and children - but if there are no more women and children around, fill up with men". We explain that, once we had noticed this, we had the idea of creating the variable "boatside" for our quantitative dataset to test this hypothesis.

(6) *Crosstabulation Way to the boat deck * Classcrew* : The way to the boat deck was different for different classes. For example, lower-class passengers had a much longer and more difficult way to the boat deck.

We explain to students that what we have effectively done in our qualitative analysis is to show the meaning and function of the quantitative variables in the context of the "social game" that was played on the *Titanic*. This social game used rules, representations, objects, actions and interactions that students had not thought of when making their initial hypotheses. The reason is that they were not familiar enough with the specifics of the "social game". Their "everyday assumptions" about the *Titanic* were incorrect (for example that first class passengers were wealthier and bribed the crew to enter the boats). This might happen for any mono-quantitativeresearch in which the researcher is not familiar enough with the case in hand.

3.5 Revisit the quantitative data

Once students have a better understanding of the social game played on the *Titanic*, and new ideas about what might be important in explaining survival probabilities, we invite them to do a second round of quantitative analysis. This round incorporates analyses that look at the specific time that individuals boarded a lifeboat and the side of the *Titanic* from where they boarded the lifeboat.

Note for SPSS: The "boatentertime" variable indicates 40.00 for 00:40 a.m. and 120.00 for 1:20 a.m.

```
*****************************************
Textbox 5: Further quantitative analysis in SPSS
*****************************************

USE ALL.
COMPUTE filter_$=(boatentertime < 130).
VARIABLE LABELS filter_$ 'boatentertime < 130 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*** (1) Crosstabs: class and boarding time
crosstabs boatentertime by classcrew
/cells count row
/statistics
/barchart.

*** (2) Crosstabs : class, sex and boarding time
crosstabs boatentertime by sex by boatside
/cells count  col
```

/statistics.

```
###########################################
Textbox 6: Further quantitative analysis in R
###########################################

data$boatside <- relevel(data$boatside, ref = "Port")


data %>%
  filter (boatside != 99)%>%
  group_by(sex, classcrew, boatside) %>%
  summarize (n = n()) %>%
  mutate(percent = n/sum(n)* 100)

data %>%
  filter (boatside != 99)%>%
  group_by(sex, classcrew, boatside) %>%
  summarize (n = n()) %>%
  mutate(percent = n/sum(n)* 100)%>%
  ggplot(aes(classcrew, percent, group = sex, fill = sex))+
  geom_col(position = position_dodge(preserve = "single"))+
  facet_wrap(~ boatside )

# (1) Calculate filling over time according to class / crew


data %>%
  filter (boatorder != 99)%>%
  group_by(sex, classcrew, boatorder) %>%
  summarize (n = n()) %>%
  mutate(percent = n/sum(n)* 100)

# (1bis) Plot lifeboat filling over time according to class / crew (Figure 1)

data %>%
  filter (boatorder != 99)%>%
  filter (boatside != 99)%>%
  group_by(boatside, classcrew1, boatorder) %>%
  summarize (n = n()) %>%
  mutate(percent = n/sum(n)* 100)%>%
  ggplot(aes(boatorder, percent, group = classcrew1, fill = classcrew1))+
  geom_col(position = position_dodge(preserve = "single"), width = 1.5)+
  facet_wrap(~ boatside )+
  labs(title = "Lifeboat filling over time according to class/crew and Port/Starboard", x = "Minutes after impact",
y = "Percent")+
  theme(legend.title = element_blank())

# (2) Plot lifeboat filling according to gender and Port/Starboard (Figure 3)

data %>%
select(boatside, sex1, boatentertime) %>%
filter(boatside != "99") %>%
```

```
filter(boatentertime < 130) %>%
group_by (boatside, sex1, boatentertime) %>%
summarize (n = n()) %>%
mutate(cum_n = cumsum(n))  %>%
ggplot(aes(x = boatentertime, y = cum_n, group = sex1, color = sex1))+
geom_line()+
facet_wrap(~ boatside)+
labs(title = "Lifeboat filling according to gender and Port/Starboard", x = "Minutes after impact", y = "Entered
lifeboat")+
scale_colour_discrete(labels = c("male", "female"))+
theme(legend.title = element_blank())
```

In their further quantitative analysis, students test the new ideas that they had during their qualitative analysis. We bring together the results in class:

(1) *Class and boarding time :* As suspected, higher-class passengers had a higher chance of boarding the lifeboats that left the *Titanic* earlier. While the qualitative material allowed us to create the hypothesis, analysis of the quantitative data leaves no doubt that this was in fact the case (Figure 2).

(2) *Class, sex and boat side:* The quantitative data confirm that rules of filling were applied differently on port and starboard. In fact, on starboard, more men boarded lifeboats than women (Figure 3). This was the case in spite of the fact that the rule "women and children first" was in fact applied, and can be explained by the fact that the rule "when there are no more women, fill with men" was applied on starboard (but not on port).
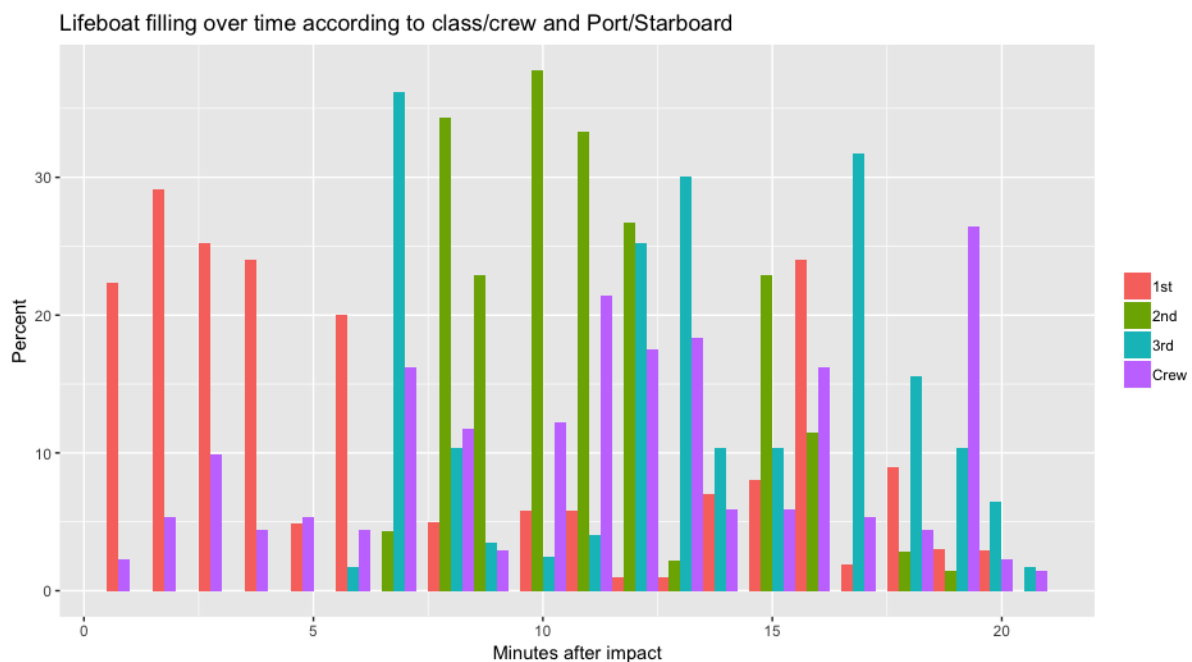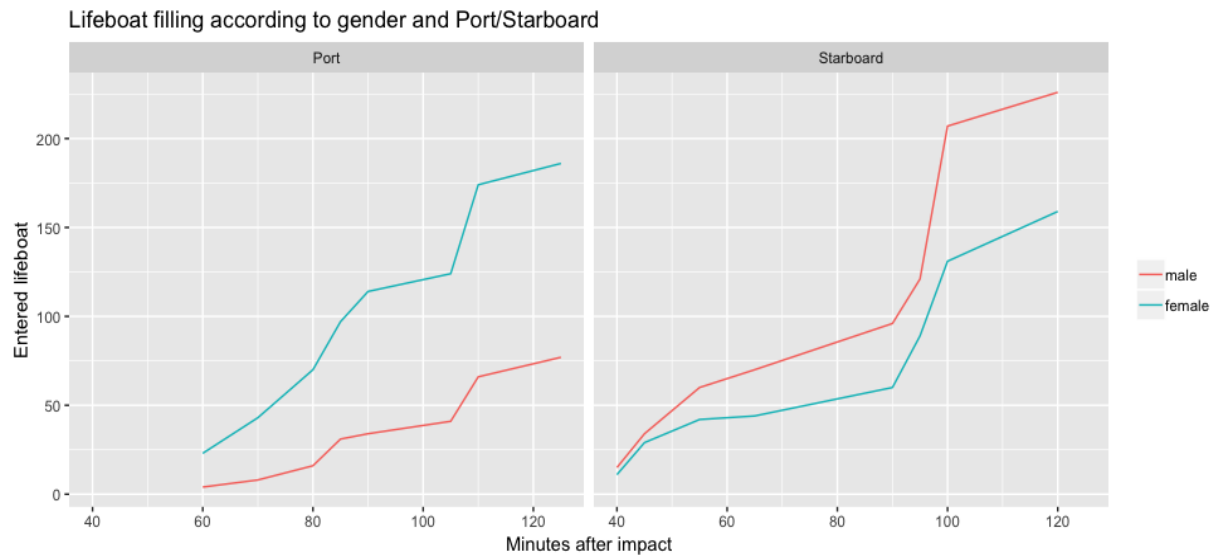
Figure 2 (this graph is only possible in "R")

Lifeboat filling over time according to class/crew and Port/Starboard

Figure 3 (this graph is only possible in "R")

Lifeboat filling according to gender and Port/Starboard

3.6 Wrapping up

In a final part, we wrap up the exercise with the students and highlight a few general points. We often do this by asking students the following questions, discussing their responses, and then making the following points (if they have not already been made by the students themselves).

(1) *Have we learned more by using both qualitativeand quantitative data than we would have had we only used one method? Are our conclusions more valid than if we had had only quantitative or only qualitative data?*
Students overwhelmingly answer in the affirmative. We then ask: why exactly? The group comes up with something like the following: with the help of the qualitative dataset, we were able to unearth information about the "social game" that we did not initially have to interpret quan-information: this information concerned the importance of the crew, the importance of rules, and the importance of time. We were able to understand the meaning and function of the quantitative variables in the context of the social game better. If students challenge this point, we remind them of their initial hypotheses that almost invariably did not include a large number of the finer points found by qualitative analysis. Conversely, we were able to quantify, correlate, and generalize many insights gleaned from the qualitative material that would have remained less convincing without quantification (frequency distributions; correlations/bivariate distributions; size and significance of effects). The conclusions we drew are called meta-inferences, which can be defined as conclusions concerning the phenomenon to be analyzed based on the analysis of qualitative and quantitative data.

(2) *Why do the datasets allow us to make meta-inferences so well?*
In the discussion, we come up with something like the following: a first point is that the data stem from the same case and the same people. This is an important point. Imagine that we had had quantitative data from the *Titanic*, but qualitative interviews with passengers of a different ship that had also sunk (say, the Costa Concordia in Italy). It would have been very difficult to

obtain useful meta-inferences. A second point is that we coded the qualitative material in such a way that it allowed us to provide a systematic description of the mechanisms leading to the outcome of interest. Imagine that we had coded for "dress code" or "metaphors". This might have been interesting – but it would not have helped us with our research question and would not have allowed us to make useful meta-inferences.

(3) *What are the limits to our analysis?*
We discuss the limits to our analysis with students. We have already mentioned this when we discussed our data, and we do not need to repeat it in detail here. Suffice to say that we talk about (a) the "survivorship-bias": clearly, only survivors can tell their story, and much of what happened to those who perished will remain unknown forever; (b) the fact that our view of the filling of the lifeboats is more reliable than our view of the way to the boat deck, since we can cross-check so many accounts with the former, but not with the latter.

(4) *How can we summarize the overall findings?*
We ask students to give a short summary of the overall results. We then give our own version that goes something like this (Stolz, Lindemann and Antonietti 2018)[vi]: Women and children survived more often than men because of the rule "Women and children first", which was the one conscious rule that officers and crew applied throughout the process. Whenever women or children were in sight, they were first allowed onto the lifeboats. However, the rule was interpreted differently on starboard (where the boats were "filled with men", once there were no more women or children in sight) and port (where only women and children, and the members of crew needed to accompany them, were allowed to board). Higher-class female passengers survived more often than lower-class female passengers, because the former arrived earlier on the boat deck, with first-class women passengers arriving earlier than second-class, and second-class earlier than third-class.
Male passengers were able to survive for reasons that changed over time. In the first phase, first-class male passengers were able to survive because of the reluctance of many women to board a lifeboat, and because they were the only men on the boat deck to "fill" the lifeboats. In the second phase, a number of lifeboats on port were lowered with a very strict rule of "women and children only", which meant that men (with the exception of male deck crew) only had a very small chance of boarding a lifeboat.
The tragedy of second-class male passengers was that they would have been present on the boat deck and in a prime position to "fill" these boats (since many crew members and third-class male passengers had not yet reached the boat deck) – but they were not allowed to do so. In the third phase, the seriousness of the situation had became obvious, and crew members and third-class male passengers seem to have been more enterprising when "filling" and "surreptitiously boarding" the lifeboats, thereby crowding out both first-class and second-class male passengers. Once in the water, younger men had an edge over older men in surviving until a lifeboat could pick them up (this was rare, however: only one woman survived in this way). The discrimination against lower classes was not a conscious policy when filling the boats. Rather, it was a combination of several mechanisms: for example, the fact that the cabins of lower-class passengers were much farther away from the boat deck, that access to the boat deck was normally denied to third-class passengers, and that there were fewer stewards to attend to them.

4. Conclusion

This paper has presented the Titanic datasets and provided a script for a three-hour exercise to teach students the general utility of mixed methods.

We have no doubt that this is a good example for teaching mixed methods, simply because we have used it successfully both in classes and workshops with beginner and advanced students for years. It never fails to engage students and it always creates moments of both puzzlement and sudden insight.

Of course, this example – as any example – has its limits. One of the most important drawbacks is the "survivorship-bias": those who perished obviously did not testify later. Much of what happened on the *Titanic* during its last hours will remain unknown forever simply because those who know did not survive to tell us. Nevertheless, there is an important lesson to be learned here, and some of the best discussions that we have with students about the *Titanic* involve this and other forms of possible bias and how to deal with it in the analysis.

A second limit is the specificity of the example that is actually seldomly found in other datasets. In this case, we have 214 individuals all telling us what happened during one and the same process, which lasted for roughly two hours. This, of course, gives us tremendous leverage and opportunities to triangulate the various testimonies. In most other studies, we do not have so many accounts all focusing on such a specific process and such a short timespan, which means that there are far fewer opportunities to triangulate and cross-check the data. Moreover, the very fact that so many people tell us what happened during those fateful hours also helps us to bring home an important point to students: namely, that there are not two different realities, one quantitative and the other qualitative. The fact that most qualitative researchers think of social reality as negotiated, constructed, fluid, context-dependent, multiple, while most quantitative researchers think of it as objective, relatively stable over time, single, and caused by various determinants, can be nicely addressed with our example. It is of course the same reality out there, independent of whether we look at it through the testimonies or the quantitative data. Everything that we observe qualitatively could also be counted and correlated; everything that we count and correlate has at its basis a socially constructed interaction that follows social rules, is embedded in social representations, and is (re)created in social action.

Appendix 1: The datasets

A1.1 The Quan-Dataset

The variables

Name	Variable Label	Values & Labels	Mis-sings	Measure-ment level	Notes
age	Age, numerical	-	11	numerical	
age_cat	Age, categories	1 = "0-14" 2 = "15-30" 3 = "31-40" 4 = "41-50" 5 = "51-60" 6 = "61+"	11	ordinal	
age51plus	Age, 51+	0 = " "	-	nominal	

		1 = "51+"			
boatentertime	Time when boarding lifeboat	-9 = "missing"	19	numerical	All people not boarding lifeboat receive time of last lifeboat (= "censored data")
boatentertime1	Time when boarding lifeboat, corrected	-9 = "missing"	19	numerical	Correction: - Assign William Murdock to Collapsible D. - Put people plucked from water boatentertime1 = 135
boatnumber	Number of lifeboat	1-10 A-D Z = "Water, not collapsible" 99 = "missing"	19	character	Z = people helped into a lifeboat from the water (but not into a collapsible)
boatorder	Order of lifeboats leaving *Titanic*	99 = "missing"	1506	ordinal	
boatside	Location of lifeboats boarded	0 = "Starboard" 1 = "Port"	1509	nominal	
group	Individual part of a group	0 = "Single" 1 = "Single w/servant" 2 = "Couple" 3 = "Couple w/children" 4 = "Couple w/servant" 5 = "Single parent w/children" 6 = "Family w/servant" 7 = "Family/Friends" 8 = "Crew Groups" 9 = "Family/Friends w/children"	-	nominal	
cc_1st	First-class passenger	0 = " " 1 "First-class passenger"	-	nominal	
cc_2nd	Second-class passenger	0 = " " 1 "Second-class passenger"	-	nominal	
cc_3rd	Third-class passenger	0 = " " 1 "Third-class passenger"	-	nominal	
cc_alacarte	A la carte crew	0 = " " 1 "A la carte crew"	-	nominal	
cc_deckcrew	Deck crew	0 = " " 1 "Deck crew"	-	nominal	
cc_enginecrew	Engine crew	0 = " " 1 "Engine crew"	-	nominal	
child	Child	0 = " " 1 "Engine crew"	-	nominal	

classcrew	Type of class; type of crew	1 = "First-class" passenger 2 = "Second-class" passenger 3 = "Third-class passenger" 4 = "A la carte Crew" 5 = "Deck Crew" 6 = "Engine Crew" 7 = "Victualling Crew"	-	nominal	
classcrew1	Type of class; type of crew (4 categories)	1 = "First-class" passenger 2 = "Second-class" passenger 3 = "Third-class passenger" 4 = "Crew"	-	nominal	
country	Country of residence	0 = "Argentina" 1 = "Austria" 2 = "Belgium" 3 = "Bosnia" 4 = "Bulgaria" 5 = "Canada" 6 = "China" 7 = "Croatia" 8 = "Cuba" 9 = "Denmark" 10 = "Egypt" 11 = "England" 12 = "Finland" 13 = "France" 14 = "Germany" 15 = "Greece" 16 = "Hungary" 17 = "India" 18 = "Ireland" 19 = "Italy" 20 = "Japan" 21 = "Lebanon" 22 = "Mexico" 23 = "Netherlands" 24 = "Northern Ireland" 25 = "Norway" 26 = "Peru" 27 = "Russia" 28 = "Scotland" 29 = "Siam" 30 = "Slovenia" 31 = "South Africa" 32 = "Spain" 33 = "Sweden" 34 = "Switzerland" 35 = "Turkey" 36 = "Uruguay" 37 = "USA" 38 = "Wales" 39 = "Yugoslavia" 40 = "Australia"	-	nominal	

country_5cat	Country of residence (5 categories)	0 = "England" 1 = "Ireland" 2 = "Sweden" 3 = "USA" 4 = "Others"	-	nominal	
country_3cat	Country of residence (3 categories)	0 = "England" 1 = "USA" 2 = "Others"	-	nominal	
du_engl	Country of residence: England	0 = " " 1 "England"	-	nominal	
du_other	Country of residence: Other	0 = " " 1 "Other"	-	nominal	
du_usa	Country of residence: USA	0 = " " 1 "USA"	-	nominal	
gr_single	Travelling alone	0 = " " 1 "Single"	-	nominal	
gr_withkids	Travelling with children	0 = " " 1 "With children"	-	nominal	
group	Travelling as part of a group	0 = "Single" 1 = "Single w/servant" 2 = "Couple" 3 = "Couple w/children" 4 = "Couple w/servant" 5 = "Single parent w/children" 6 = "Family w/servant" 7 = "Family/friends" 8 = "Crew groups" 9 = "Family/friends w/children"	-	nominal	
ID	ID		-	nominal	
lived	Survived yes/no	0 = "Perished" 1 = "Survived"	-	nominal	
name	Name		-	nominal	
sex	Sex	0 = "Male" 1 = "Female"	-	nominal	
testimony	Testimony available	0 = "No" 1 = "Yes"	-	nominal	

The dataset and syntax
The datasets can be downloaded on www.mixedmethodstitanic.wordpress.com.

Formats syntax
.sps (SPSS)
.R (R)

Formats dataset
Titanic_Mixed.sav (SPSS)
Titanic_Mixed.csv (comma separated values, for use in R)

A1.2 The Qual-Dataset

The documents
The coded documents can be downloaded on www.mixedmethodstitanic.wordpress.com.

The coding scheme
The coding scheme can be downloaded on www.mixedmethodstitanic.wordpress.com.

Formats
.mx12 (MAXQDA, coded documents)
.pdf (coding scheme and documents separately)

Appendix 2: Reproducing the results in the published papers

A2.1 The R-code

The complete R code can be downloaded on www.mixedmethodstitanic.wordpress.com.

References

Bazeley, Patrizia. 2003. "Teaching Mixed Methods." *Qualitative Research Journal* (3 (Special Issue)):117-26.

Bellocco, Rino, and Sara Algeri. 2013. "Goodness-of-fit tests for categorical data." *The Stata Journal* 13(2):356-65.

Bryman, Alan. 2008. "Why do Researchers Integrate/Combine/Mesh/Blend/Mix/Merge/Fuse Quantitative and Qualitative Research?" Pp. 87-100 in *Advances in Mixed Methods Research. Theories and Applications*, edited by Manfred Max (Ed.) Bergman. Los Angeles: SAGE.

Creswell, John W., Abbas Tashakkori, Ken D. Jensen, and Kathy L. Shapley. 2003. "Teaching Mixed Methods Research: Practices, Dilemmas, and Challenges." Pp. 619-38 in *Handbook of Mixed Methods in Social & Behavioral Research*, edited by Abbas Tashakkori and Charles Teddlie. Thousand Oakes: Sage.

Kelle, Udo. 2001. "Sociological Explanations between Micro and Macro and the Integration of Qualitative and Quantitative Methods [43 paragraphs]." *Forum: Qualitative Social Research [On-line Journal]. Available at: http://qualitative-research.net/fqs/fqs-eng.htm [Date of access: 10.04.2008]* 2(1).

—. 2007. *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung. Theoretische Grundlagen und methodologische Konzepte*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kohler, Ulrich, and Frauke Kreuter. 2017. *Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung. 5. Auflage*. Oldenbourg: De Gruyter.

Landau, Sabine, and Brian S. Everitt. 2004. *A Handbook of Statistical Analyses using SPSS*. London: Chapman & Hall/CRC.

Maxwell, Joseph A., and Kavita Mittapalli. 2010. "Realism as a Stance for Mixed Methods Research." in *Mixed Methods in Social & Behavioral Research. Second Edition*, edited by Abbas Tashakkori and Charles Teddlie. Thousand Oaks: Sage.

Stolz, Jörg. 2016. "Opening the black box. How the study of social mechanisms can benefit from the use of explanatory mixed methods." *Analyse & Kritik* 37(1 (Social Mechanisms)):257-85.

—. 2017. "Institutional, Alternative, Distance, and Secular. Four Types of (Un)Belief and their Gods." *Nordic Journal of Religion and Society* 30(1):4-23.

Stolz, Jörg, Judith Könemann, Mallory Schneuwly Purdie, Thomas Englberger, and Michael Krüggeler. 2016. *(Un)Believing in modern society. Religion, spirituality, and religious-secular competition*. London: Routledge.

Stolz, Jörg, and Anaïd Lindemann. 2019. "The Titanic Game : Introducing game heuristics to mixed methods research (forthcoming)." *Journal of Mixed Methods Research*.

Stolz, Jörg, Anaïd Lindemann, and Jean-Philippe Antonietti. 2018. "Sociological Explanation and mixed methods: the example of the *Titanic*." *Quality and Quantity* 53(3):1623-43.

Strauss, Anselm. 2003(1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambrdige University Press.

---

[i] We thank David A. Savage for providing us with the initial dataset.

[ii] This is measured with a proxy: the time when the boat actually left the Titanic (and which is reasonably well know).

[iii] https://www.encyclopedia-titanica.org.

[iv] http://www.titanicinquiry.org/about.php.

[v] There is some overlap between this section describing the datasets and validity issues and the same description in Stolz, Lindemann & Antonietti (2018), "Sociological Explanation and mixed methods: the example of the *Titanic*". *Quality and Quantity* 53(3):1623-43. We believe that this is acceptable and do not claim originality for this part of the paper. What is needed here is consistency over the different papers, and not originality.

[vi] This summary contains some findings that were not seen in the exercises.