

Analysing survey data in Public Health

Chris Shaw

12 May 2016

Introduction

Surveys are an essential part of policy making in Public Health. Limited resources have to be allocated to deliver the maximum outcomes. Conducting a survey across a wide spectrum of public health professionals is a valuable way to gather prevailing information and prioritise policy areas to address.

This paper examines effective ways to analyse and present survey data to support the decision making process about areas to concentrate on.

Raw survey data

The appendix shows a simulated survey with 10 questions and 19 responses. The raw data is captured in an excel spreadsheet in a tabulated manner. The topics are the column headings. Each row corresponds to one response. The participant's topic preferences are indicated by a number - 1 being the highest preference. Only three preferences are allowed per response in strict order.

The analysis will work on any number of topic columns, responses or preference choices.

```
xlsfile="surveydata.xlsx"
```

```
calculate_scores <- function(number_to_select=4) {  
  # Get the data from excel and flatten into a data set that can be manipulated more easily  
  response_data <- read.xlsx(xlsfile,1)  
  topic_data <- read.xlsx(xlsfile,2)  
  
  respondent <- names(response_data)[1]  
  survey<-melt(response_data, id.vars = respondent, variable.name="topic",  
              na.rm=TRUE, value.name="priority")  
  survey$RespondentID <- as.factor(survey$RespondentID)  
  number_respondents=nrow(response_data)  
  
  # Calculate the score for each response  
  lowest_priority=max(survey$priority)  
  survey <- survey %>% mutate(score=(lowest_priority + 1 - priority)^2)  
  
  # Calculate the total score per topic  
  topic_scores <- survey %>% group_by(topic) %>% summarise(topic_score=sum(score)) %>%  
    arrange(topic_score)  
  
  # Put a 1 in the weight column for the last number of topics determined by  
  # number_to_select variable. Other topics have weight zero. This selects the topics  
  # which are the most popular  
  topic_scores$weight <- c(rep(0, nrow(topic_scores)-number_to_select),
```

```

        rep(1,number_to_select))

# Add the topic descriptions
topic_scores <- merge(topic_scores, topic_data, by="topic")
topic_scores <- arrange(topic_scores, topic_score)

# Mix the weights into the survey data
survey <- merge(survey, topic_scores, by="topic")

# Caluculate the satisfaction for each respondent against the selected most
# popular topics
survey <- survey %>% mutate(satisfaction=score^2*weight)

resp <- survey %>% group_by(RespondentID) %>%
  summarise(satisfaction=sum(satisfaction)) %>%
  arrange(satisfaction)

resp<<-resp
topic_scores<<-topic_scores
survey<<-survey
}

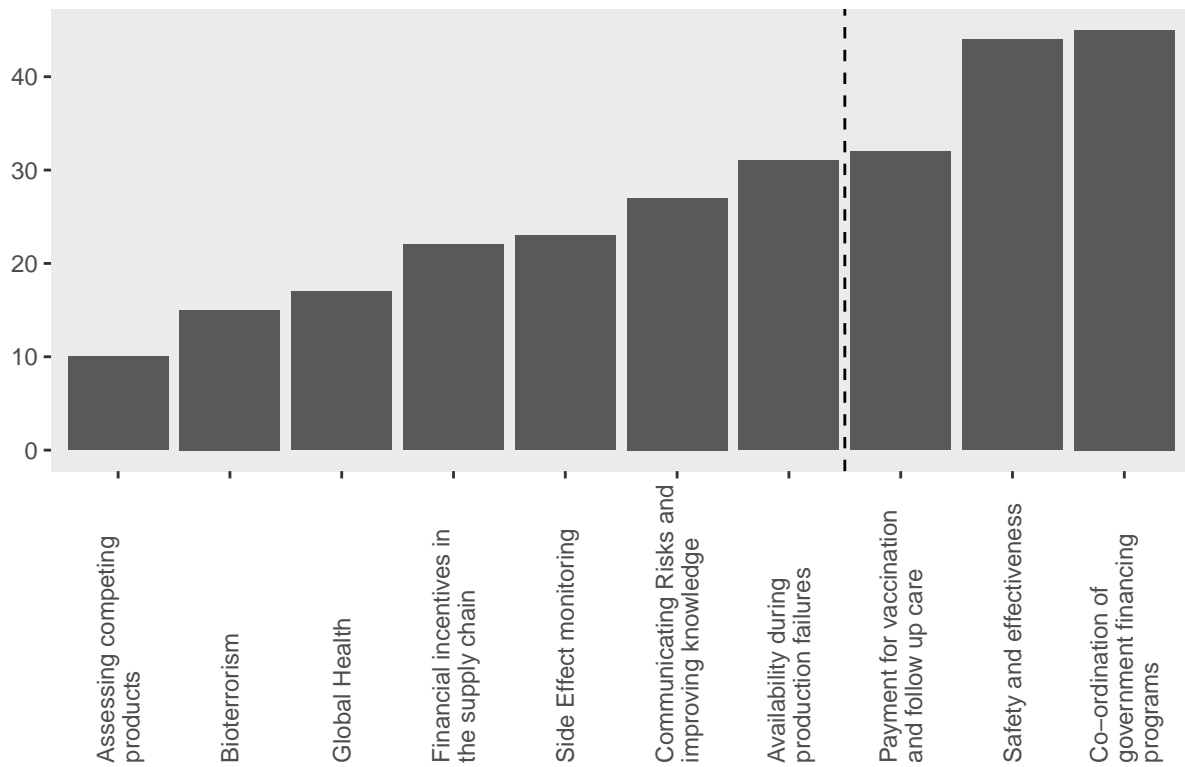
calculate_scores(3)
# Create the y axis labels with the subject of each topic
ylabels <- str_wrap(as.vector(topic_scores$subject), width = 25)
names(ylabels)<-topic_scores$topic

plot_topic_score <- function(number_to_select){
  ggplot(topic_scores, aes(x=topic, y=topic_score)) +
    geom_bar(stat="identity") +
    scale_x_discrete(limits=topic_scores$topic,
                     labels=ylabels) +
    ylab("") + xlab("") +
    theme(panel.grid.minor=element_blank(),
          panel.grid.major=element_blank(),
          axis.text.x = element_text(angle = 90, hjust=0, vjust=0.5)) +
    geom_vline(aes(colour=4), xintercept = nrow(topic_scores)- number_to_select +0.5,
               linetype="dashed")
}

plot_topic_score(3)+ggtitle("Three most popular topics across all respondents")

```

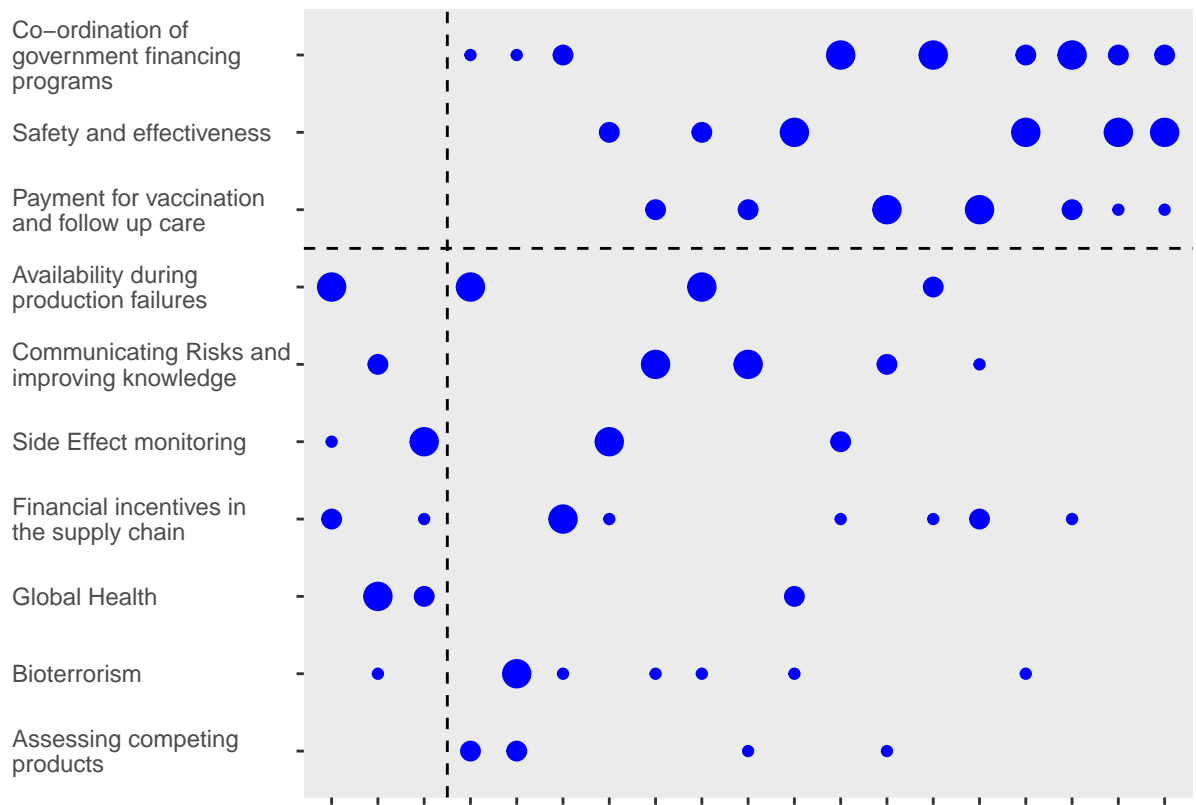
Three most popular topics across all respondents



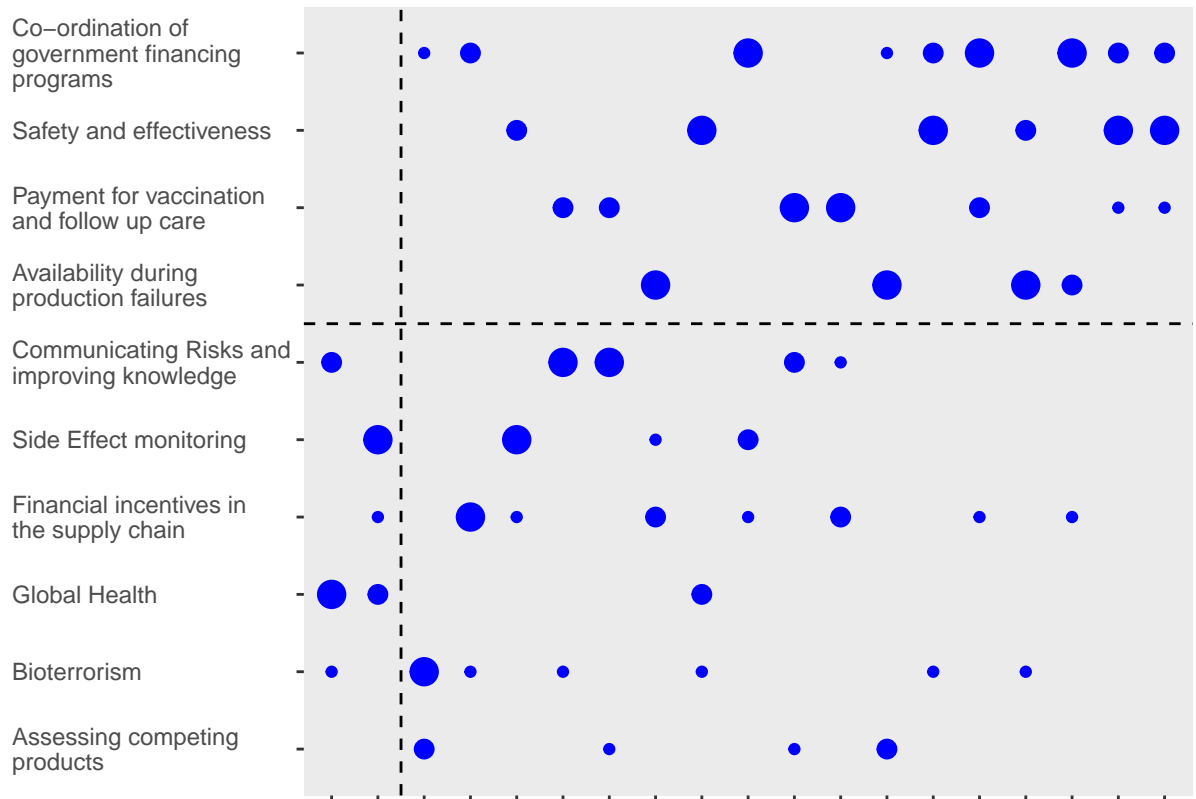
```
plot_responses <- function (number_to_select) {

  ggplot(survey, aes(x=RespondentID, y=topic)) +
    geom_point(size=sqrt(survey$score)*1.5, colour="blue") +
    scale_y_discrete(limits=topic_scores$topic,
                     labels=ylabels) +
    ylab("") + xlab("") +
    scale_x_discrete(limits=resp$RespondentID) +
    theme(panel.grid.minor=element_blank(),
          panel.grid.major=element_blank(),
          axis.text.x = element_blank(),
          axis.text.y = element_text(hjust=0)) +
    geom_hline(yintercept = nrow(topic_scores) - number_to_select + 0.5,
               linetype="dashed") +
    geom_vline(xintercept = nrow(resp[resp$satisfaction==0,])+0.5,
               linetype="dashed")
}

plot_responses(3)
```



```
calculate_scores(4)
plot_responses(4)
```



Appendix

```
response_data[is.na(response_data)] <- ""
kable(response_data, align=rep("c", ncol(response_data)))
```

RespondentID	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10
1				1	2			3		
2		3			1	2				
3	1		2		3					
4		1			2			3		
5			2		3			1		
6	1	2					3			
7				1				3		2
8		3		2			1			
9						2		3	1	
10		3			1		2			
11								3	2	1
12	1			2				3		
13			3			1			2	
14	2	3			1					
15		2				1			3	
16				1	2	3				
17		3					1			2
18			3			2			1	
19				1	2	3				

```
kable(topic_data)
```

topic	subject
Topic1	Availability during production failures
Topic2	Financial incentives in the supply chain
Topic3	Assessing competing products
Topic4	Safety and effectiveness
Topic5	Co-ordination of government financing programs
Topic6	Payment for vaccination and follow up care
Topic7	Side Effect monitoring
Topic8	Bioterrorism
Topic9	Communicating Risks and improving knowledge
Topic10	Global Health