Chapter 9

Mapping Study Analysis

9.1 Analysis of publication details 102
9.2 Classification analysis 103
9.3 Automated content analysis 106
9.4 Clusters, gaps, and models 110

This chapter is the first of three consecutive chapters dedicated to data synthesis. We have split this topic into three chapters because the three different types of systematic review (quantitative, qualitative and mapping study) require very different procedures. Data synthesis is also one of the tasks that many software engineering researchers have identified as least well addressed by current guidelines, see Cruzes & Dybå (2011*b*) and Guzmán, Lampasona, Seaman & Rombach (2014).

Chapter 10 discusses qualitative synthesis which is suitable for systematic reviews of qualitative primary studies as well as systematic reviews of quantitative primary studies that are unsuitable for meta-analysis. Chapter 11 describes statistical methods used to synthesise primary studies that report quantitative comparisons of different software engineering techniques.

This chapter discusses the analysis methods used for mapping study reviews. We address mapping studies first because the analysis methods used to summarise results from mapping studies are generally quite straightforward but provide some insight into the problems experienced synthesising results from systematic reviews. The analysis of mapping study results is relatively simple because the data extracted from each primary study in a mapping study are much less detailed than the data extracted from primary studies in systematic reviews. However, more complex analyses based on text mining can help identify clusters of similar studies either to validate the study inclusion and exclusion process or identify subsets of studies for more detailed analysis.

The examples of data analysis presented in this chapter were analysed using the R statistical language. We strongly recommend using R since:

- It is free and open source.
- There are numerous ancillary packages developed by statistical experts.
- There are many textbooks describing R. We personally recommend *R in Action* (Kabacoff 2011).
- It provides flexible programming facilities.

9.1 Analysis of publication details

Many research questions can be answered by analysing publication details, such as:

- Author name and affiliation
- Publication date
- Publication type
- Publication source

Such data are usually analysed as simple tables of counts such as the number of publications per author, or per country of affiliation, or simple trend-based graphics such as the number of publications per year. For example, Mair, Shepperd & Jørgensen (2005) analysed empirical cost estimation studies to identify and investigate the data sets. They presented:

- A table showing the number of studies in each journal that was used in the search process.
- A line plot of the number of studies grouped in three-year periods.

When a mapping study is performed as part of a PhD thesis, a research student needs to know which papers are the most influential in their field, and to have read and understood them. Using a general indexing system such as Scopus, Web of Science, or Google Scholar, it is easy to discover how many papers have cited each primary study, which is a good way of identifying such papers.

Another type of analysis that can be useful when dealing with a large number of primary studies is to look for author networks, that is, groups of authors that have collaborated to produce a number of primary studies. This information can be used with classification data to identify whether groups of authors concentrate on specific topics, problems or methods. For example, in the context of a meta-analysis of machine learning defect prediction methods, Shepperd, Bowes & Hall (2014) used the author group as a moderator factor in their analysis. Surprisingly, it proved to be the most important moderator factor, accounting for 31% of the variation among studies. Although the primary studies all compared different prediction methods, prediction method as a factor accounted for only 1.3% of the variation.

It may also be useful to analyse the cross-references within a set of primary studies to look for clusters of studies and isolated studies. It is also worth checking whether the isolated studies should really have been excluded. Furthermore, analysis of the combined set of included and exclude studies can be used to check that inclusion and exclusion criteria have been used consistently. For example, if publication detail analysis shows that some excluded papers are among a cluster of included papers, they should be re-assessed for possible inclusion.

9.2 Classification analysis

The more interesting research questions are usually based on classifying the primary studies and may concern issues such as:

- Identifying the existing research approaches and/or concrete techniques used in a topic area and cross-referencing between the approach taken and the relevant primary studies.
- Identifying the experimental methods used in empirically-based studies.
- Mapping approaches and techniques to the overall software engineering process or to specific steps in a specific software engineering task.

For example, Mair et al. (2005), provide a large number of diagrams[1] relevant to describing and categorizing software engineering datasets. The diagrams show:

- The number and percentage of datasets that were available or unavailable to researchers.
- The number of datasets collected in three-year time periods.
- The size of the datasets in terms of number of projects.
- The dataset size, in terms of numbers of features (attributes).
- The frequency of dataset usage (that is, the number of studies that used each dataset).

Mair et al. did not specify *a priori* research questions for their mapping study, but, as experts in the topic area, they provided analyses of great interest to cost estimation researchers.

In another example, Elberzhager, Rosbach, Münch & Eschbach (2012) in a mapping study on methods to reduce test effort, asked the question:

> "What are existing approaches for reducing effort when applying testing techniques, and how can they be classified?"

---

[1] The diagrams are labelled histograms but are actually frequency diagrams.

To address the question they extracted keywords from the abstracts of 144 primary studies and then looked for additional keywords by reading the introductions and conclusions of the primary studies. This allowed them to identify five groups of testing methods (specifically Test automation, Prediction, Test input reduction, Quality Assurance before testing, Test strategy). They then tabulated the number and percentage or papers in each category, but they also provided a narrative description of each approach.

To be helpful to readers, data displays should allow the reader to track papers to the categories that describe them. This is usually done by presenting all the extracted information in data matrices (which need to be either published in the review reported or included in ancillary information available to readers), but can also be incorporated into data analysis displays. For example, Elberzhager et al. looked at papers in each of the categories in more detail and identified more detailed subcategories. For test automation they identified different phases of code automation and displayed the information using a horizontal bar chart with the identifiers of each paper printed beside the relevant horizontal row. An artificial example of a horizontal bar chart is shown in Figure 9.1. This was obtained from the data and the code snippet shown in Figure 9.2.

Petersen et al. (2008) suggest the use of bubble plots to visualize relationships among categorical variables. An example showing the structure of a bubble plot can be seen in Figure 9.3.

A bubble plot assumes we have three categorical variables and want to plot two of the variables (X-Variable 1 and X-Variable 2) against the third (Y-Variable). The relationships are shown by the number of studies that share a specific X-variable category and a specific Y category variable. For example, in Figure 9.3, 15 studies exhibit both Y-Variable category 4 and X-Variable 1 Category 2. In this case, the value 15 means that 21.3% of the papers that have been categorized according to X-Variable 1. A bubble plot does not assume that every study is categorized against each variable (e.g., some studies may not exhibit any of the categories associated with a X-variable and other studies may exhibit several different categories of the same X-variable), nor does it display any direct relationship between the X-variables.

Bubble plots can be produced manually using a drawing package. Alternatively, R supports bubble plots of two-variables but to produce the bubble plot shown in Figure 9.3, the data must be organized as shown in Table 9.1, the X-Variable 1 has 5 categories which are mapped to the values xpos=-5, -4,...,-1 while X-Variable 2 has 6 categories mapped to the values xpos=1,2,...,6. The Y-variable has 5 categories mapped to the values yvar=1,2,..5. *val* identifies the number of primary studies that have the specific X-category and Y-category. The values of *xtpos* and *ytpos* identify the (x,y) co-ordinates on the bubble plot where the percentages associated with *val* should be printed so they are displaced from the bubble. The R code used should be based on the snippet shown in Figure 9.4.

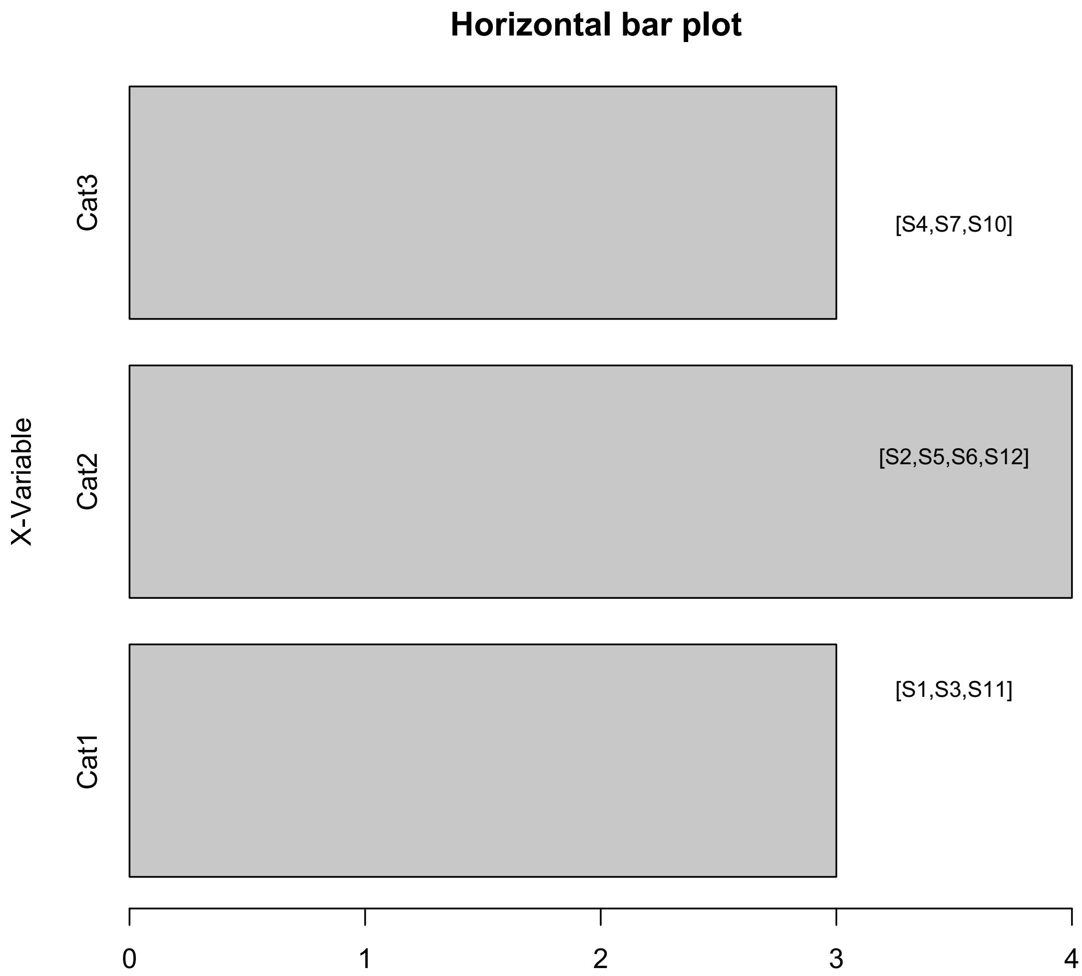

FIGURE 9.1: Example of a horizontal bar chart including study IDs.

A limitation of classification methods used in mapping studies is that although the set of categories relating to a specific feature or characteristic may appear to be mutually exclusive, primary studies are often more complex. For example, mapping study analysis often uses the classification of study types Wieringa, Maiden, Mead & Rolland (2006) proposed to classify requirements engineering papers. The categories include: Problem investigation, Solution design, Solution Validation, Solution selection, Solution implementation, Implementation evaluation. However, in practice, a paper discussing a “Solution design” will often include a section demonstrating the feasibility of the proposed solution which would be an example of “Solution Validation”. This means that such a paper should be classified in both categories. It is often clear from bubble plots or tabular displays that the total number of classified papers is greater than the number of primary studies but it is not clear from bubble plots which papers exhibit multiple categories. Furthermore, if researchers categorise a paper in terms of their personal opinion of the ‘main’

```
#Input the categorical data
XVar=c("Cat1","Cat2","Cat1","Cat3","Cat2","Cat2","Cat3","Cat3","Cat1","Cat2")
#Identify the number of studies in each category
counts=table(XVar)
#Creat a horizontal bar plot
barplot(counts, main="Horizontal bar plot", ylab="X-Variable",horiz=T)
#Add Text identifying the studies to the bar chart
text(3.5,1,"[S1,S3,S11]",cex=0.8)
text(3.5,2,"[S2,S5,S6,S12]",cex=0.8)
text(3.5,3,"[S4,S7,S10]",cex=0.8)
```

FIGURE 9.2: Bar chart code snippet.

goal of the paper, then some categories may be artificially underrepresented. Multiple classifications of primary studies or underrepresentation of certain categories make it more difficult to understand the implications of the reported frequency counts.

9.3 Automated content analysis

Recently, several researchers have suggested the use of text mining and associated visualization methods to analyse mapping study data, see Felizardo, Nakagawa, Feitosa, Minghim & Maldonado (2010) and Felizardo et al. (2012). These techniques can be used for analysing citations among papers as described in Section 9.1, however, in this section, we describe their use for content analysis. Content analysis and text mining can be used to:

- Check inclusion and exclusion decisions during primary study selection.
- Identify clusters of studies that might be suitable for more detailed analysis as a set of related studies.

Text mining and visualization require specialist tools (see Chapter 13). Felizardo et al. (2012) used the following process for content mapping using the *Revis* tool:

1. *Text preprocessing* is used to structure and clean the data. They used text from the title, abstract and keywords only. In addition, the text is analysed to create a vector of terms (words) present in the text which are weighted based on *term frequency-inverse document frequency measurement* which involves weighting words:
    - in direct proportion to its frequency in a specific primary study, but

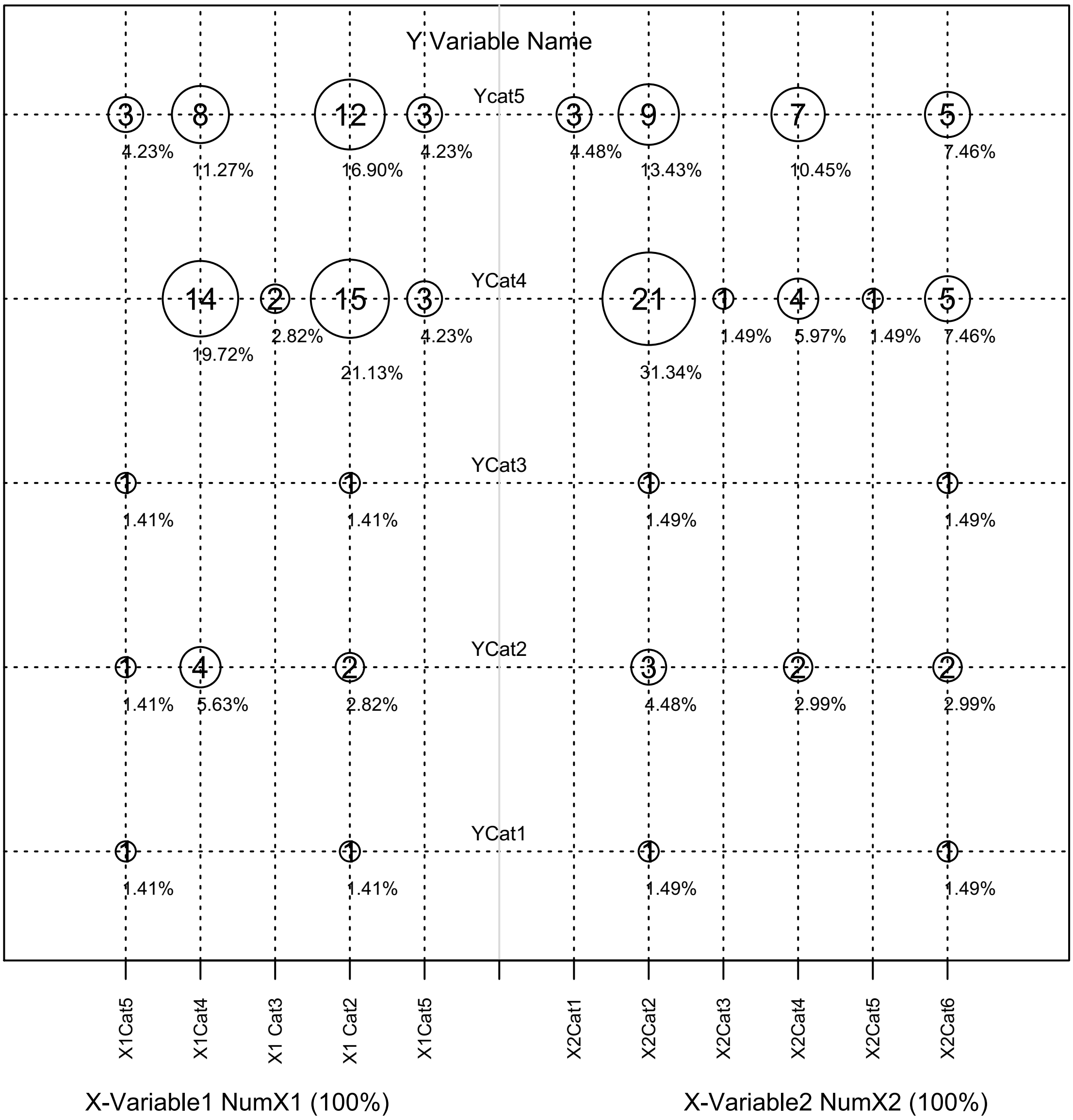

FIGURE 9.3: Example of a bubble plot showing the structure.

- in inverse proportion to its frequency in the other studies.

2. *Similar Calculation* which uses the vector of weighted words to calculate (dis)similarity among primary studies. Felizardo et al. used a method based on cosines: $distance(i, j) = 1 - cos(\bar{x}_i, \bar{x}_j)$ where $\bar{x}_i$ and $\bar{x}_j$ are vectors of weights for the ith and jth primary studies.

3. *Projection* which maps the m-dimensional vectors onto 2 or three dimensions that can be represented visually.

TABLE 9.1: Bubble Plot Data

xvar	yvar	val	xtpos	ytpos	percent
−5	5	3	−4.7	4.8	4.23%
−5	3	1	−4.7	2.8	1.41%
−5	2	1	−4.7	1.8	1.41%
−5	1	1	−4.7	0.8	1.41%
−3	4	2	−2.7	3.8	2.82%
−4	2	4	−3.7	1.8	5.63%
−4	4	14	−3.7	3.7	19.72%
−4	5	8	−3.7	4.7	11.27%
−2	1	1	−1.7	0.8	1.41%
−2	2	2	−1.7	1.8	2.82%
−2	3	1	−1.7	2.8	1.41%
−2	4	15	−1.7	3.6	21.13%
−2	5	12	−1.7	4.7	16.9%
−1	4	3	−0.7	3.8	4.23%
−1	5	3	−0.7	4.8	4.23%
2	1	1	2.3	0.8	1.49%
2	2	3	2.3	1.8	4.48%
2	3	1	2.3	2.8	1.49%
2	4	21	2.3	3.6	31.34%
2	5	9	2.3	4.7	13.43%
1	5	3	1.3	4.8	4.48%
3	4	1	3.3	3.8	1.49%
4	2	2	4.3	1.8	2.99%
4	4	4	4.3	3.8	5.97%
4	5	7	4.3	4.7	10.45%
5	4	1	5.3	3.8	1.49%
6	1	1	6.3	0.8	1.49%
6	2	2	6.3	1.8	2.99%
6	3	1	6.3	2.8	1.49%
6	4	5	6.3	3.8	7.46%
6	5	5	6.3	4.8	7.46%

The projection maps can be colour coded to show whether studies that appear similar based on content analysis have received the same inclusion and exclusions decision. They point out that clusters can be of two types:

1. Pure clusters where all primary studies received the same in(ex)clusion decision. Such clusters do not need to be reassessed.

2. Mixed clusters where some primary studies were included and some excluded. Felizardo et al. (2012) suggest reassessing any primary studies

```
bplot=read.table("filename.txt",head=T)
summary(bplot)
attach(bplot)

r=val/100
ny=c(1,2,3,4,5,6)

#Draws the circles - the spaces in the subtitle are intentional
symbols(xvar,yvar,circle=sqrt(r/pi), inches=.25,
xlab="",ylab="",xaxt="n",yaxt="n", sub="X-Variable1 NumX1 (100%)
X-Variable2 NumX2 (100%)",cex.sub=.8)

#Adds the numbers to the circles
text(xvar,yvar,val)
#Adds the grid lines
abline(h=c(1,2,3,4,5),lty=3)
abline(v=c(-5,-4,-3,-2,-1),lty=3)
abline(v=c(1,2,3,4,5,6),lty=3)

#Adds a central y-line
abline(v=c(0),lty=1,col="Yellow")

# Defines labels for the x-axis
labx=c("X1Cat5","X1Cat4","X1 Cat3","X1
Cat2","X1Cat5","","X2Cat1","X2Cat2","X2Cat3","X2Cat4","X2Cat5", "X2Cat6")
!
tckx=c(-5,-4,-3,-2,-1,0,1,2,3,4,5,6)

#Adds labels tick marks on the x-axis
axis(1,at=tckx,labels=labx,cex.axis=.6,las=2)

#Defines labels for the y-axis

laby=c("Ycat5", "YCat4","YCat3","YCat2", "YCat1")

#Specifies position of Y labels
nlab=c(5.1,4.1,3.1,2.1,1.1)

#Adds Y labels to plot
text(0,nlab,laby,cex=.65)
#Adds name to y-axis
text(0,5.4,"Y Variable Name",cex=.8)

#Adds the offset percentage information
text(xtpos,ytpos,percent,cex=0.6)
```

FIGURE 9.4: Bubble plot code snippet.

found in such clusters. If only one or two studies were in(ex)cluded, they needed to be reassessed in order to determine whether they should be reclassified to conform with majority decision.

In addition, isolated points need to be reassessed if they have been included.

9.4 Clusters, gaps, and models

We defined the main goals of mapping studies as finding clusters of studies suitable for more detailed studies and identifying gaps where more research is needed (see Chapter 3). In order to identify useful clusters and meaningful gaps, it is necessary to have some theoretical model of the mapping study topic against which the primary studies can be assessed. This may be a generic classification scheme such as that proposed by Wieringa et al. (2006), but it could be a classification scheme derived from an existing model of the software engineering processes addressed by the topic (for example the three layer model of cloud engineering), or of the way in which the existing software processes would be changed by the topic (for example the way test-before changes the overall testing process). We note however, that, although the identification of a large number of papers in a particular category is a strong indicator of a cluster, the absence of primary studies particularly in two-way tables or bubble plots does not necessarily imply a gap in the literature. It might mean that the specific combination of categories is either not meaningful or not important. To be identified as a topic suitable for further research, a gap needs a convincing explanation of why further primary research is likely to be important.

It is also possible that a mapping study might lead to the development of a model of the topic area, as an outcome of reading and classifying the literature. At the moment, this is an underutilised approach in mapping studies, but, as we point out in Section 10.4, Popay, Roberts, Sowden, Petticrew, Arai, Rodgers, Britten, Roen & Duffy (2006) suggest that the starting point of a narrative synthesis of a systematic review should be a model of the topic of interest. So if a mapping study is intended to be the starting point of a systematic review, it may be useful to consider whether its results can be represented as a model of the topic area, used to organise the primary studies.