Thesis Proposal

Visualization methods for genealogical and RNA-sequencing datasets

Lindsay Rutter

Program of Study Committee:

Dianne Cook, Major Professor
Amy Toth, Major Professor
Heike Hofmann
Daniel Nettleton
James Reecy

May 16, 2016

Contents

Chapter 1

Introduction

1.1 Background to data visualization

Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah.

1.2 Problems to be addressed

Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah.

1.3 Overview of thesis research

Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah.

Chapter 2

Visualization methods for genealogical datasets

2.1 Abstract

In this chapter, we introduce ggenealogy ([3]), a developing R software package that provides tools for searching through genealogical data, generating basic statistics on their graphical structures using parent and child connections, and displaying the results. The package allows users to draw the genealogy in relation to variables related to the nodes, and to determine and display the shortest path distances between the nodes. Production of pairwise distance matrices and genealogical diagrams constrained on generation are also available in the visualization toolkit. We are testing the tools on a dataset with milestone cultivars of soybean varieties ([1]) as well as on a web-based database of the academic genealogy of mathematicians ([2]). The software package has been available on the Comprehensive R Archive Network since March 2015.

Chapter 3

Stock syntax

3.1   Database structure

Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah Blah blah blah. As an example, the learning outcomes for `Topic 03` are provided in List 1 below. As seen in Figure 3.1.

Figure 3.1: Caption blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah

Now, we will run the REFER TO SECTION .

Table 3.1: Topic numbers and descriptions

Number	Description
01	Data
02	Descriptive Statistics for a Single Categorical Variable
03	Descriptive Statistics for a Single Quantitative Variable

List 1: Learning outcomes for Topic 03

A. Use standardizing to determine how many standard deviations an observation is away from the mean value.

B. Use z-scores to compare observations for different quantitative variables.

C. Explain how standardizing affects the shape, center, and variability of the distribution of a quantitative variable.

Example question title: **T16.A.A.04-1.1.MC.1**

The absolute pathway to the `extdata` directory on your local computer can be determined by typing the following command into the **R** console:

```
system.file("inst/extdata/", package = "ePort")
```

```
keyHTM = system.file("inst/extdata/KeyFiles/Topic06.Questions.htm", package =
  "ePort")

refineKey(keyHTM)

keyPath = gsub("htm$", "txt", keyHTM)
```

```
dataPath = system.file("inst/extdata/DataFiles/Topic06/Topic06.A.csv", package =
  "ePort")

rewriteData(dataPath)

loPath = system.file("inst/extdata/LOFiles/Topic06.Outcomes.txt", package = "ePort")

outPath = system.file("inst/extdata/OutputFiles", package = "ePort")

makeReport(keyFile = keyPath, dataFile = dataPath, loFile = loPath, outFile =
  outPath)
```

```
merged = subsetData(mergedData, dataTable)
makeReport(outFile = outPath, unit = 2, reportType = "crossSecUnit", className =
  "Eng444", repeatLowScore = 70)
```

- One topic for one section - short version ("secTopicShort")

- One topic for one section - long version ("secTopicLong")

- One topic comparing multiple sections - short version ("crossSecTopicShort")

- One topic comparing multiple sections - long version ("crossSecTopicLong")

- One unit (group of topics) for one section ("secUnit")

- One unit (group of topics) comparing multiple sections ("crossSecUnit")

Bibliography

[1] T Hymowitz, CA Newell, and SG Carmer. *Pedigrees of Soybean Cultivars Released in the United States and Canada.* International Soybean Series, College of Agriculture, University of Illinois at Urbana-Champaign, Urbana, IL, 1977.

[2] *The Mathematics Genealogy Project*, 2010. Archived Web Site. Retrieved from the Library of Congress, Accessed on March 6, 2015.

[3] Lindsay Rutter, Susan Vanderplas, and Di Cook. *ggenealogy: Visualization Tools for Genealogical Data*, 2015. R package version 0.1.0.