The effects of interaction on linear projections of multivariate data

Nick, Di, Kim

Abstract

abstract text here

Introduction

Hypothesis

Does the finer control afforded by the manual tour improve the ability of the analyst to understand the importance of variables contributing to the structure?

Experimental design

Factors

Each of `N` Participants were randomly split into 3 even groups. The first group saw only a single static linar projection, that of first two principal componets of the data. The second group watched a 30 second loop of Grand tour. The final group was allowed to control an interactive manual tour for their duration.

Block treatments

Each participant performed each of 4 block treatments in random order. The blocks consisted of determing the the dimensionality of the dataset, `p`, the number of the clusters, `n`, the number of important variables, `d`, and if there existed an significant covariance, `s`.

Randomization & replication

Paricipants were randomly assigned to one of 3 factors deciding which visual methed they recieved. The blocks were performed in a random order for each participant. Within each block, pacticipants performed 4 replications, answering the block question for each of the 4 datasets in a random order before proceeding to the next block.

Response & measures

Each block was introduced and demonstrated directly preceeding each block. During this introductory question each participant was shown the visual for their factor with a writen description of the block and how to descern it with the same toy data set. Participants recieved exactly 2 minutes to study/explore each repitition's projection before answering a question regarding it. Answers came in the form of a numeric input for three blocks - namely, dimensionality, clusters, and imporant variables (`p`, `n`, and `d` respectivly). For the remaining block, covariance `s`, a checkmark box was provided for each varible. Pacticipants we instructed to mark all variables, if any, that were highly correlated. None of the data sets contained more than a single group of highly-correlated variables.

After