User Perceptions of Sound in Simple Linear Regression Diagnostics

Jessica M. Orth[1]
Advisor: Kate Cowles[1]

[1]Department of Statistics and Actuarial Science, The University of Iowa, 241 Schaeffer Hall Iowa City, IA 52242-1409

March 2014

Outline

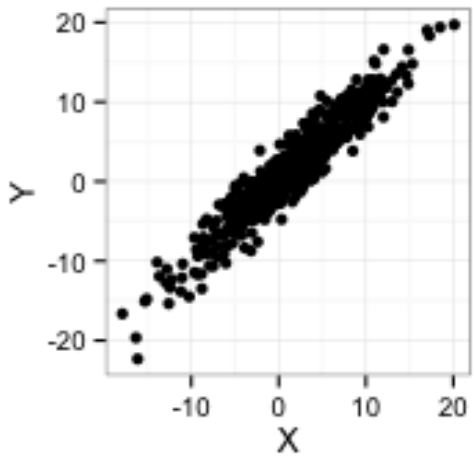
Outline

Outline

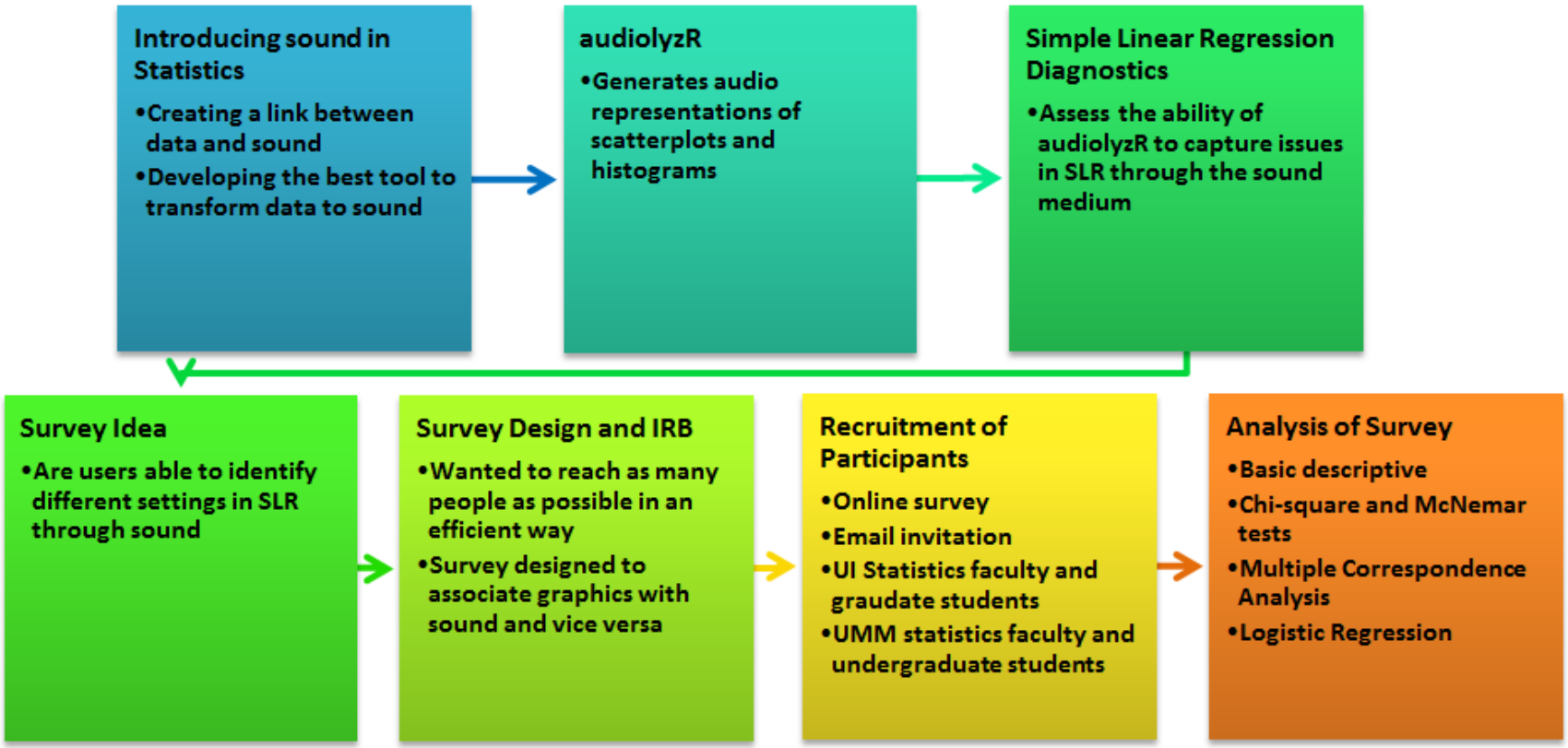
Outline

Outline

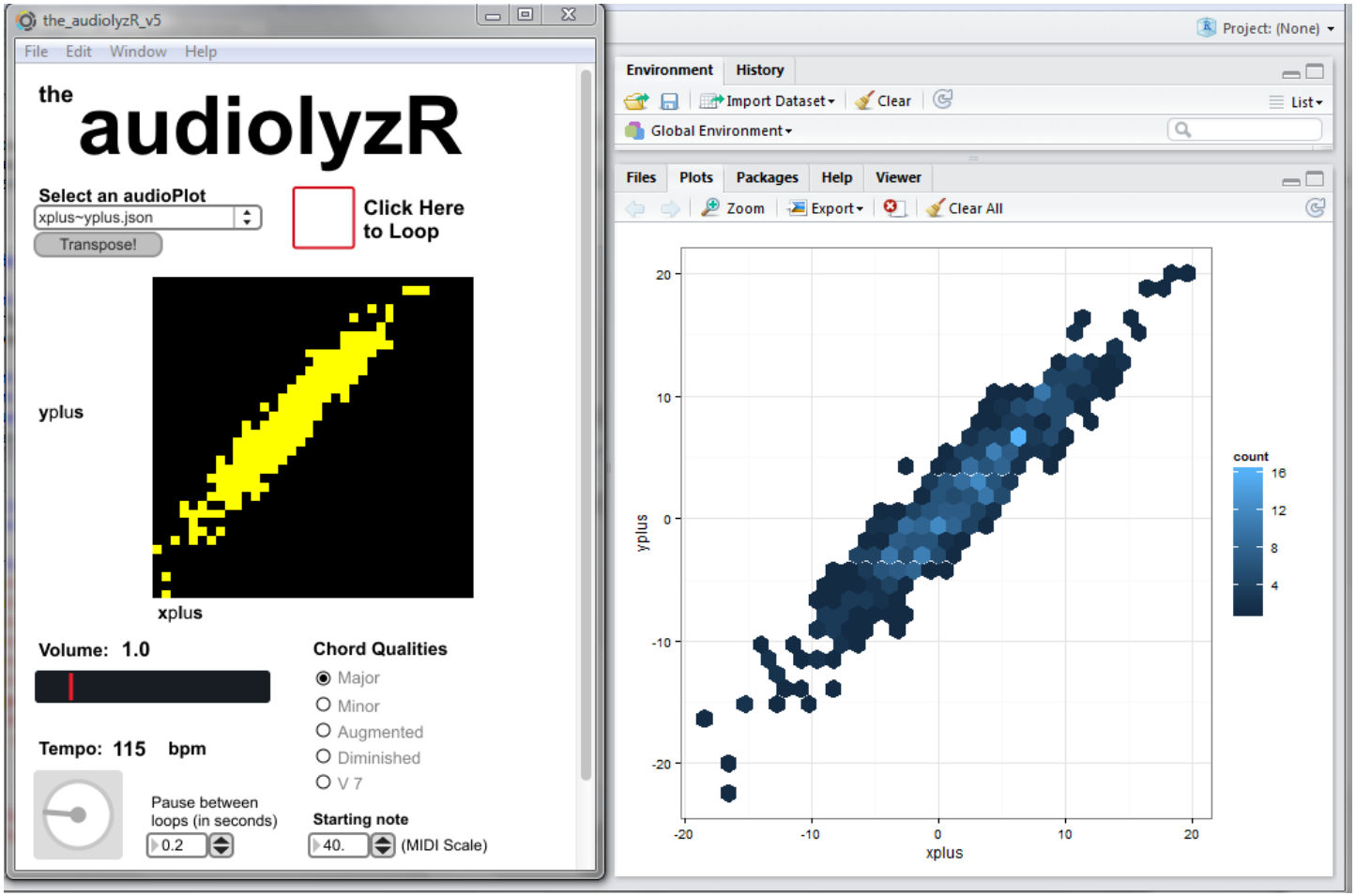
Introduction



Approach


Introducing sound in Statistics
•Creating a link between data and sound
•Developing the best tool to transform data to sound
audiolyzR
•Generates audio representations of scatterplots and histograms
Simple Linear Regression Diagnostics
•Assess the ability of audiolyzR to capture issues in SLR through the sound medium
Survey Idea
•Are users able to identify different settings in SLR through sound
Survey Design and IRB
•Wanted to reach as many people as possible in an efficient way
•Survey designed to associate graphics with sound and vice versa
Recruitment of Participants
•Online survey
•Email invitation
•UI Statistics faculty and graudate students
•UMM statistics faculty and undergraduate students
Analysis of Survey
•Basic descriptive
•Chi-square and McNemar tests
•Multiple Correspondence Analysis
•Logistic Regression

audiolyzR


the_audiolyzR_v5
File Edit Window Help
the audiolyzR
Select an audioPlot
xplus~yplus.json
Transpose!
Click Here to Loop
yplus
xplus
Volume: 1.0
Chord Qualities
Major
Minor
Augmented
Diminished
V 7
Tempo: 115 bpm
Pause between loops (in seconds)
0.2
Starting note
40.
(MIDI Scale)
Project: (None)
Environment
History
Import Dataset
Clear
List
Global Environment
Files
Plots
Packages
Help
Viewer
Zoom
Export
Clear All
count
16
12
8
4
20
10
0
-10
-20
yplus
xplus

Survey Design

Survey Implementation

Qualtrics

- Online survey design

Institutional Review Board

- Approved November 26, 2013

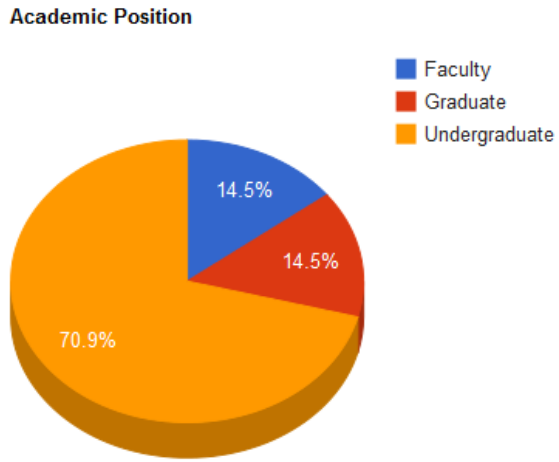
Survey Launch

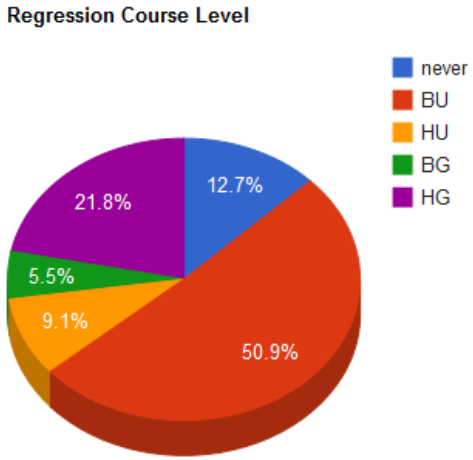
- December 6, 2013
- Statistics faculty and students at
  - The University of Iowa and University of Minnesota Morris

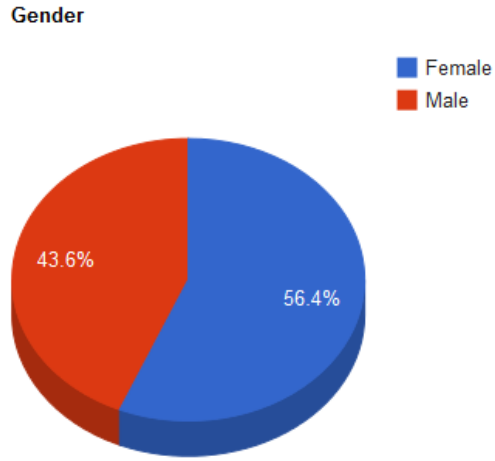
Survey Results

- 101 responses
- 55 completed surveys

Background information







Analysis of control questions

Control Question	Overall	Position	Level	Gender
Association	1.00			
Variance	0.84 (0.74, 0.94)	Faculty: 1.00 Graduate: 1.00 Undergraduate: 0.77 (0.1729)	BU: 0.86 HG: 1.00 Other: 0.67 (0.05372)*	Female: 0.87 Male: 0.79 (0.4817)
Functional	0.85 (0.76, 0.94)	Faculty: 1.00 Graduate: 1.00 Undergraduate: 0.79 (0.2115)	BU: 0.86 HG: 1.00 Other: 0.73 (0.1159)	Female: 0.87 Male: 0.83 (0.7176)
Normality	0.65 (0.52, 0.78)	Faculty: 0.88 Graduate: 0.75 Undergraduate: 0.59 (0.3673)	Never: 0.71 BU: 0.57 HG: 0.83 Other: 0.63 (0.5852)	Female: 0.71 Male: 0.58 (0.3973)

Table : Proportion of correct responses for overall participants and background characteristics for control questions. Numbers correspond to proportion correct (95% confidence) for the overall column and proportion correct (p-value for Fisher's Exact Test for Count Data) for the remaining columns

Analysis of control questions

Control Question	Overall	Position	Level	Gender
Transformation	0.75 (0.64, 0.86)	Faculty: 1.00 Graduate: 1.00 Undergraduate: 0.64 (0.01899)*	Never: 0.71 BU: 0.61 HG: 1.00 Other: 0.88 (0.06413)	Female: 0.71 Male: 0.79 (0.5471)
Outlier	0.87 (0.78, 0.96)	Faculty: 0.63 Graduate: 0.75 Undergraduate: 0.95 (0.02385)*	Never: 1.00 BU: 0.96 HG: 0.58 Other: 0.88 (0.01481)*	Female: 0.94 Male: 0.79 (0.2197)

Table : Proportion of correct responses for overall participants and background characteristics for control questions. Numbers correspond to proportion correct (95% confidence) for the overall column and proportion correct (p-value for Fisher's Exact Test for Count Data) for the remaining columns

Analysis of sound questions

Sound to Graphics Question	Overall	Position	Level	Gender
Association	0.56 (0.43, 0.69)*	Faculty: 0.75 Graduate: 0.88 Undergraduate: 0.46 (0.06459)	BU: 0.57 HG: 0.83 Other: 0.33 (0.06014)	Female: 0.55 Male: 0.58 (1.00)
Functional	0.78 (0.67, 0.89)	Faculty: 1.00 Graduate: 0.75 Undergraduate: 0.74 (0.344)		Female: 0.77 Male: 0.79 (1.00)
Variance	0.44 (0.31, 0.57)*		BU: 0.46 HG: 0.45 Other: 0.40 (1.00)	Female: 0.52 Male: 0.35 (0.2738)

Table : Proportion of correct responses for overall participants and background characteristics for sound to graphics questions. Numbers correspond to proportion correct (95% confidence) for the overall column and proportion correct (p-value for Fisher's Exact Test for Count Data) for the remaining columns. Empty cells are a consequence of anonymity in a small sample size.

Analysis of sound questions

Graphics to Sound Question	Overall	Position	Level	Gender
Outlier	0.61 (0.48, 0.74)*		BU: 0.75 HG: 0.55 Other: 0.40 (0.1125)	Female: 0.65 Male: 0.57 (0.584)
Transformation	0.57 (0.44, 0.70)*		BU: 0.54 HG: 0.67 Other: 0.57 (0.7036)	Female: 0.67 Male: 0.46 (0.1687)

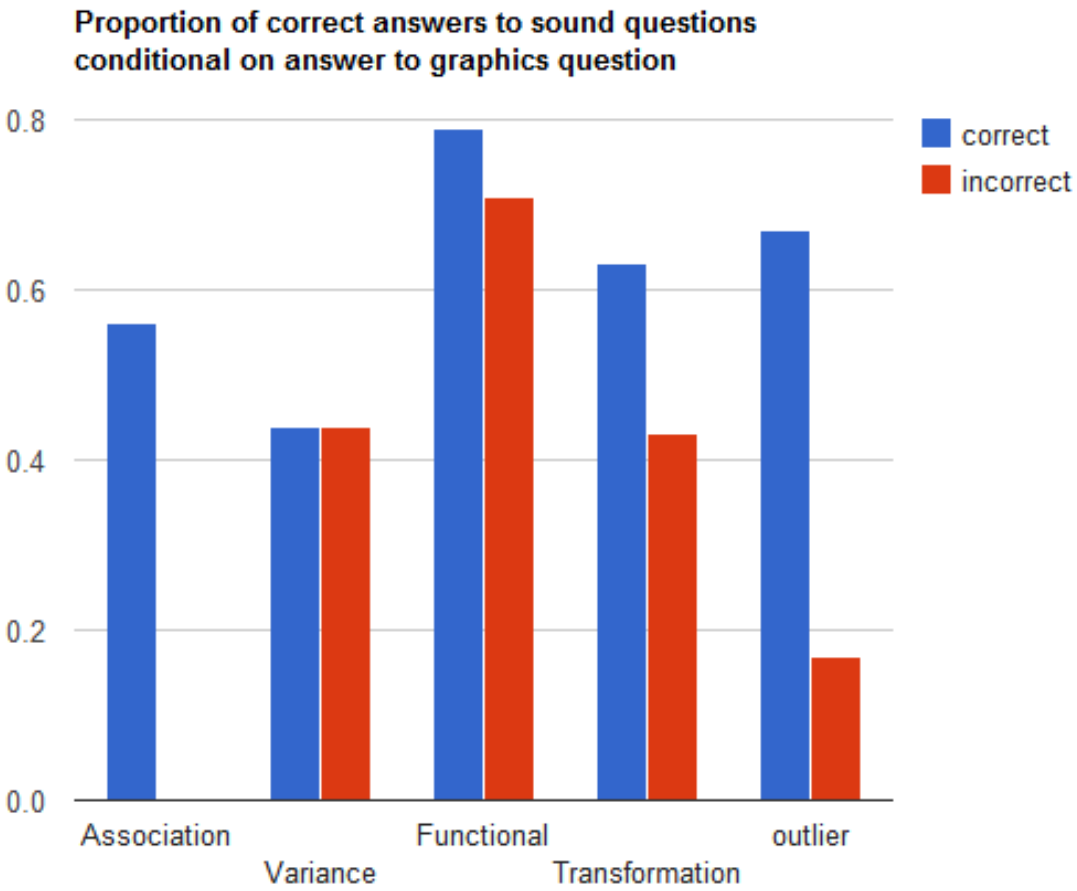
Table : Proportion of correct responses for overall participants and background characteristics for graphics to sound questions. Numbers correspond to proportion correct (95% confidence) for the overall column and proportion correct (p-value for Fisher's Exact Test for Count Data) for the remaining columns. Empty cells are a consequence of anonymity in a small sample size.

McNemar's test

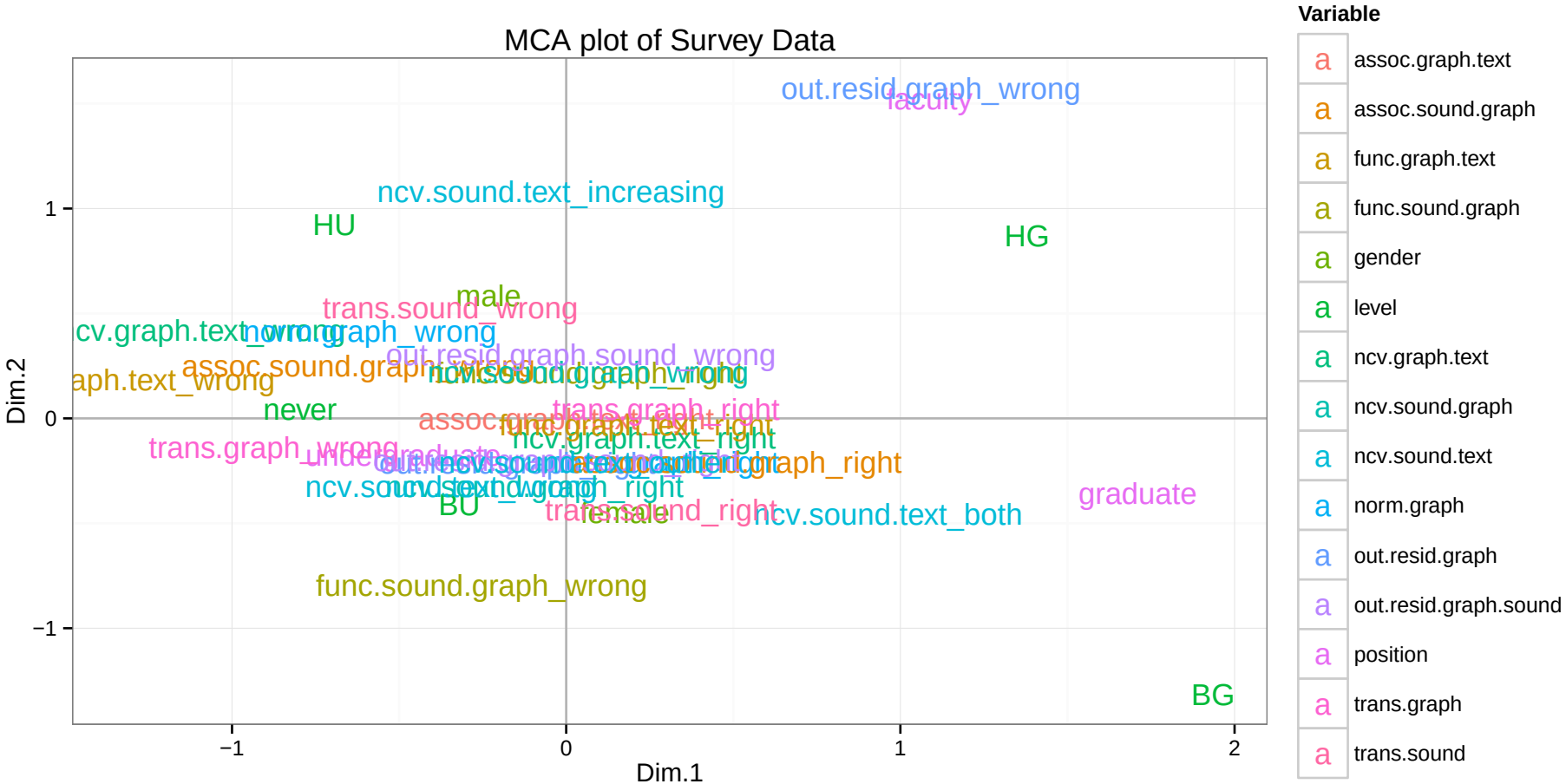
Question Pairs (graphic vs. sound)	McNemar's p-value	Proportion of correct answers to sound question conditional on a correct answer to graphics question	Proportion of correct answers to sound question conditional on an incorrect answer to graphics question
Association	2.668e-06*	0.56	
Variance	0.0002041*	0.44	0.44
Functional	0.3017	0.79	0.71
Transformation	0.08086	0.63	0.43
Outlier	0.000685*	0.67	0.17

Table : p-values for McNemar's test of independence in question pairs (graphic vs. sound) and the conditional proportion of correct answers to sound questions given the response to the corresponding graphics question

McNemar's test



Multiple correspondence analysis



Logistic regression

$$y_i \leftarrow Y_i \ indep \sim binom(n_i, \mu_i),$$
$$logit\mu_i = \beta_0 + \beta_1 outlier.graphic_i, i = 1, \ldots, 55,$$
$$where \ \mu_i = P(outlier.sound \ correct | outlier.graphic)$$

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6931	0.3062	2.26	0.0236
outlier.graphic.wrong	-2.3026	1.1374	-2.02	0.0429

Table : Estimates for logistic regression model

$$logit(\hat{\mu}(outlier.graphic)) = 0.6931 - 2.3026 * outlier.graphic_{wrong}$$

Impact of the survey

- The survey was received with much enthusiasm from participants, who indicated a strong interest in the idea of introducing sound to Statistics
- Overall theme of evaluations was that the idea of introducing sound to Statistics was very interesting, with a high potential for future applications
- Evaluations indicated a strong interest in applying sound to the Statistics classroom to help students who learn better by listening and also those who are visually impaired

Conclusions

- Users are able to correctly identify different diagnostic plots in simple linear regression through the sound medium
- Certain regression concepts were more challenging to identify through sound: identification of outliers and non-constant variance
- Users are able to connect sound with graphical displays and vice versa
- `audiolyzR` used in the simple linear regression setting proved to be effective in using sound to detect regression concepts

Future Work

- Explore the meaning of changing chord quality, volume, tempo, starting note, and delay time in `audiolyzR`
- Extend the use of sound in Statistics beyond the simple linear regression setting to statistical inference, simulation studies, and Bayesian analyses
- Introduce sound to the use of copulas and directional dependence structures and angular correlation
- Reach a new audience with a tool that allows for innovative ways to hear the stories behind the data
- Create a method of teaching statistics that will help those with visual disabilities

Acknowledgments

- Professor Cowles
- Survey Participants
- University of Minnesota, Morris Statistics Faculty
- Professor Sungur
- Matt Bognar

References

1 Herv Abdi and Dominique Valentin. Multiple correspondence analysis. *Encyclopedia of Measurements and Statistics*, pages 651-657, 2007.

2 Ethan Brown and Nick Bearman. Listening to uncertainty: Information that sings, *Significance*, 9(5):14-17, 2012.

3 B. Everitt and G. Dunn. *Applied Multivariate Data Analysis*. Hodder Arnold UK, United Kingdom, 2001.

4 J. Fox, *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications, Inc, USA, 2008.

5 J. Fox and S. Weisberg. *An R Companion to Applied Regression*. SAGE, Inc., USA, 2011.

6 Markus Gesmann and Diego de Castillo. googlevis: Interface between r and the google visualization api. *The R Journal*, 3(2):40-44, December 2011.

References

7 Francois Husson, Julie Josse, Sebastien Le, and Jeremy Mazet. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*, 2013. R package version 1.25.

8 P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman and Hall, United Kingdom, 1989.

9 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

10 RStudio and Inc. *shiny: Web Application Framework for R*, 2013, R package version 0.8.0.

11 Eric Stone and Jessie Garrison. *audiolyzR: Give your data a listen*, 2013. R package version 0.4-9.

12 W.N. Venables and B.D. Ripley. *Modern Applied Statistis with S*. Springer, New Yourk, fourth edition, 2002. ISBN 0-387-95457-0.