MANCHESTER 1824
The University of Manchester

The Effects of Visual and Design Features on the Perception of Correlation in Scatterplots

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Science and Engineering

2024

Gabriel Strain
Department of Computer Science

Contents

Contents **2**

List of figures **6**

List of tables **7**

List of publications **8**

Abstract **9**

Lay abstract **10**

Declaration of originality **11**

Copyright statement **12**

Acknowledgements **13**

1 Introduction **14**
1.1 Research Motivation 14
1.2 Contributions 14
1.3 Included Publications 14
1.4 Overview of Thesis 14

2 Literature Review **15**
2.1 Data Visualisation: A Brief History 15
2.2 Measuring Relatedness 15
2.3 Conceptions of Correlation 15
2.4 Visualising Correlation 15
2.4.1 History 15
2.4.2 Present Landscape 15
2.4.3 Scatterplots 15
2.5 Correlation Perception 15
2.6 Correlation Cognition 15
2.7 Underestimation: What's Really Going On? 15
2.8 Underestimation: Potential Consequences 15
2.9 Data Visualisation Literacy 15
2.10 Objectives and Contributions 15

3 General Methodology **16**
3.1 Introduction . . . 16
3.2 Experimental Methods . . . 17
3.2.1 Experimental Design . . . 17
3.2.2 Tools for Testing . . . 17
3.2.3 Creating Stimuli . . . 17
3.2.4 Recruitment . . . 17
3.3 Analytical Methods . . . 17
3.3.1 Linear Mixed-Effects Models . . . 17
3.3.2 Advantages Over Aggregate-Level Statistical Tests . . . 17
3.3.3 Model Construction . . . 17
3.3.4 Effects Sizes . . . 17
3.3.5 Reporting Analyses . . . 17
3.4 Computational Methods . . . 17
3.4.1 Executable Reporting . . . 17
3.4.2 Containerised Environments . . . 17
3.5 Reproducibility In This Thesis . . . 17
3.5.1 Sharing Data and Code . . . 17
3.5.2 Executable Papers and Docker Containers . . . 17
3.5.3 Experimental Resources . . . 17
3.6 Conclusion . . . 17

4 Adjusting the Opacities of Scatterplot Points Can Affect Correlation Estimates **18**
4.1 Abstract . . . 19
4.2 Introduction . . . 19
4.2.1 Overview . . . 19
4.3 Related Work . . . 19
4.3.1 Transparency, Contrast, Opacity, and Formal Definitions . . . 19
4.3.2 Effects of Point Opacity on Correlation Estimation . . . 19
4.4 Experiment 1: Uniform Opacity Adjustments . . . 19
4.4.1 Introduction . . . 19
4.4.2 Methods . . . 19
4.4.3 Analysis . . . 19
4.4.4 Discussion . . . 19
4.5 Experiment 2: Spatially-Dependent Opacity Adjustments . . . 19
4.5.1 Introduction . . . 19
4.5.2 Methods . . . 19
4.5.3 Analysis . . . 19
4.5.4 Discussion . . . 19
4.6 General Discussion . . . 19

5 Adjusting the Sizes of Scatterplot Points Can Correct for a Historic Correlation Underestimation Bias **20**

5.1 Abstract . . . 20
5.2 Overview . . . 20
5.3 Related Work . . . 20
5.3.1 Size and Perception . . . 20
5.3.2 Scatterplot Point Size and Correlation Perception . . . 20
5.4 Experiment: Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots . . . 20
5.4.1 Introduction . . . 20
5.4.2 Methods . . . 20
5.4.3 Analysis . . . 20
5.4.4 Discussion . . . 20
5.5 General Discussion . . . 20

6 Interactions of Opacity and Size Adjustments 21
6.1 Abstract . . . 21
6.2 Overview . . . 21
6.3 Related Work . . . 21
6.3.1 Size and Opacity . . . 21
6.4 Experiment: Adjusting Point Size and Opacity Together . . . 21
6.4.1 Introduction . . . 21
6.4.2 Methods . . . 21
6.4.3 Analysis . . . 21
6.4.4 Discussion . . . 21
6.5 General Discussion . . . 21

7 Visual Features Affecting Perceptual Estimates Also Affect Beliefs About Correlations 22
7.1 Abstract . . . 23
7.2 Overview . . . 23
7.3 Related Work . . . 23
7.3.1 From Perception to Cognition . . . 23
7.3.2 From Cognition to Belief . . . 23
7.4 Pre-Study: Investigating Beliefs About Relatedness Statements . . . 23
7.4.1 Introduction . . . 23
7.4.2 Methods . . . 23
7.4.3 Analysis . . . 23
7.4.4 Discussion . . . 23
7.5 Experiment: Potential for Belief Change Using Atypical Scatterplots . . . 23
7.5.1 Introduction . . . 23
7.5.2 Methods . . . 23
7.5.3 Analysis . . . 23
7.5.4 Discussion . . . 23
7.6 General Discussion . . . 23

8 Conclusion **24**
8.1 Main Findings 24
8.2 Relationship to Prior Work 24
8.3 Reproducibility 24
8.4 Contributions 24
8.5 Implications 24
8.5.1 For Design 24
8.5.2 For Society 24
8.6 Limitations 24
8.7 Future Directions 24
8.8 Closing Remarks 24

Appendices **25**

A First appendix **26**

Word count: 1000

List of figures

List of tables

List of publications

Publications go here.

Abstract

put abstract here

Lay abstract

This is lay abstract text.

Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made *only* in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations/`) and in The University's policy on Presentation of Theses.

Acknowledgements

Acknowledgements go here.

Chapter 1

Introduction

1.1 Research Motivation

1.2 Contributions

1.3 Included Publications

The research described in chapters 4, 5, 6, and 7 in this thesis is adapted from earlier publications, the last of which under review as of writing. To avoid repetition, information and discussion that would be repeated has been consolidated into the literature review and general methodology chapters. *Gabriel Strain* is the primary author of all included papers.

- *The Effects of Contrast on Correlation Perception in Scatterplots* [**strain_2023**] is reproduced in Chapter 4. Sections 4.4.2, 4.5.2, 4.4.3, 4.5.3, 4.4.4, 4.5.4, and4.6 contain minimally altered parts of the published article.
- *Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots* [**strain_2024**] is reproduced in Chapter 5. Sections 5.4.2, 5.4.3, 5.4.4, and 5.5 contain minimally altered parts of the published article.
- *Effects of Point Size and Opacity Adjustments in Scatterplots* [**strain_2024b**] is reproduced in Chapter 6. Sections 6.4.2, 6.4.3, 6.4.4, and 6.5 contain minimally altered parts of the published article.
- *Effects of Alternative Scatterplot Designs on Belief* (*under review*) is reproduced in Chapter 7. Sections 7.4, 7.5.2, 7.5.3, 7.5.4, and 7.6 contain minimally altered parts of the published article.

1.4 Overview of Thesis

Chapter 2

Literature Review

2.1 Data Visualisation: A Brief History

2.2 Measuring Relatedness

2.3 Conceptions of Correlation

2.4 Visualising Correlation

2.4.1 History

2.4.2 Present Landscape

2.4.3 Scatterplots

2.5 Correlation Perception

2.6 Correlation Cognition

2.7 Underestimation: What's Really Going On?

2.8 Underestimation: Potential Consequences

2.9 Data Visualisation Literacy

2.10 Objectives and Contributions

Chapter 3

General Methodology

3.1 Introduction

In this chapter we describe our research methodologies. Chapters 4, 5, and 6 share most aspects of experimental method, while the experiment described in chapter 7 differs substantially. Throughout this chapter, the reader should assume that we are referring to the entire body of experimental work this thesis describes. Methods that differ regarding the final experiment in chapter 7 are detailed along the way. In this chapter, we discuss our experimental designs, the tools we use to build and run our experiments, our approach to statistical analyses, and the computational methods and practices we employed particularly with regards to reproducibility and open science.

3.2 Experimental Methods

3.2.1 Experimental Design

3.2.2 Tools for Testing

3.2.3 Creating Stimuli

3.2.4 Recruitment

3.3 Analytical Methods

3.3.1 Linear Mixed-Effects Models

3.3.2 Advantages Over Aggregate-Level Statistical Tests

3.3.3 Model Construction

3.3.4 Effects Sizes

3.3.5 Reporting Analyses

3.4 Computational Methods

3.4.1 Executable Reporting

3.4.2 Containerised Environments

3.5 Reproducibility In This Thesis

3.5.1 Sharing Data and Code

3.5.2 Executable Papers and Docker Containers

3.5.3 Experimental Resources

3.6 Conclusion

Chapter 4

Adjusting the Opacities of Scatterplot Points Can Affect Correlation Estimates

4.1 Abstract

4.2 Introduction

4.2.1 Overview

4.3 Related Work

4.3.1 Transparency, Contrast, Opacity, and Formal Definitions

4.3.2 Effects of Point Opacity on Correlation Estimation

4.4 Experiment 1: Uniform Opacity Adjustments

4.4.1 Introduction

4.4.2 Methods

4.4.3 Analysis

4.4.4 Discussion

4.5 Experiment 2: Spatially-Dependent Opacity Adjustments

4.5.1 Introduction

4.5.2 Methods

4.5.3 Analysis

4.5.4 Discussion

Chapter 5

Adjusting the Sizes of Scatterplot Points Can Correct for a Historic Correlation Underestimation Bias

5.1 Abstract

5.2 Overview

5.3 Related Work

5.3.1 Size and Perception

5.3.2 Scatterplot Point Size and Correlation Perception

5.4 Experiment: Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots

5.4.1 Introduction

5.4.2 Methods

5.4.3 Analysis

5.4.4 Discussion

5.5 General Discussion

Chapter 6

Interactions of Opacity and Size Adjustments

6.1 Abstract

6.2 Overview

6.3 Related Work

6.3.1 Size and Opacity

6.4 Experiment: Adjusting Point Size and Opacity Together

6.4.1 Introduction

6.4.2 Methods

6.4.3 Analysis

6.4.4 Discussion

6.5 General Discussion

Chapter 7

Visual Features Affecting Perceptual Estimates Also Affect Beliefs About Correlations

7.1 Abstract

7.2 Overview

7.3 Related Work

7.3.1 From Perception to Cognition

7.3.2 From Cognition to Belief

7.4 Pre-Study: Investigating Beliefs About Relatedness Statements

7.4.1 Introduction

7.4.2 Methods

7.4.3 Analysis

7.4.4 Discussion

7.5 Experiment: Potential for Belief Change Using Atypical Scatterplots

7.5.1 Introduction

7.5.2 Methods

7.5.3 Analysis

7.5.4 Discussion

Chapter 8

Conclusion

8.1 Main Findings

8.2 Relationship to Prior Work

8.3 Reproducibility

8.4 Contributions

8.5 Implications

8.5.1 For Design

8.5.2 For Society

8.6 Limitations

8.7 Future Directions

8.8 Closing Remarks

Appendices

Appendix A

First appendix

A.1 Section in Appendix