

# **COVID-19 impact statement**

Optional. Delete this section if not needed. This is COVID-19 impact text.

# **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

<b>Contents</b>	<b>2</b>
<b>List of figures</b>	<b>6</b>
<b>List of tables</b>	<b>7</b>
<b>List of publications</b>	<b>8</b>
<b>Abstract</b>	<b>9</b>
<b>Lay abstract</b>	<b>10</b>
<b>Declaration of originality</b>	<b>11</b>
<b>Copyright statement</b>	<b>12</b>
<b>Acknowledgements</b>	<b>13</b>
<b>1 Introduction</b>	<b>14</b>
1.1 Research Motivation . . . . .	14
1.2 Contributions . . . . .	14
1.3 Overview of Thesis . . . . .	14
<b>2 Literature Review</b>	<b>15</b>
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
<b>3 General Methodology</b>	<b>16</b>

3.1	Introduction . . . . .	16
3.2	Experimental Methods . . . . .	16
3.2.1	Experimental Design . . . . .	16
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
<b>4</b>	<b>Adjusting the Opacities of Scatterplot Points Can Affect Correlation Estimates</b>	<b>18</b>
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	Experiments . . . . .	19
4.4.1	Experiment 1: Uniform Opacity Adjustments . . . . .	19
4.4.2	Experiment 2: Spatially-Dependent Opacity Adjustments . . . . .	19
4.5	General Discussion . . . . .	19
<b>5</b>	<b>Adjusting the Sizes of Scatterplot Points Can Correct for a Historic Correlation Underestimation Bias</b>	<b>20</b>
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
<b>6</b>	<b>Interactions of Opacity and Size Adjustments</b>	<b>21</b>
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
<b>7</b>	<b>Scatterplots Optimised for Correlation Perception Affect Beliefs About Correlations and Attitudes Towards Trivial Topics</b>	<b>22</b>
7.1	Abstract . . . . .	22
7.2	Overview . . . . .	22
7.3	Related Work . . . . .	22
7.3.1	From Perception to Cognition . . . . .	22
7.3.2	From Cognition to Belief . . . . .	22
7.4	Experiment: I don't know what is going to happen yet . . . . .	22
7.4.1	Introduction . . . . .	22
7.4.2	Methods . . . . .	22
7.4.3	Analysis . . . . .	22
7.4.4	Discussion . . . . .	22
7.5	General Discussion . . . . .	22
<b>8</b>	<b>Conclusion</b>	<b>23</b>
8.1	Main Findings . . . . .	23
8.2	Relationship to Prior Work . . . . .	23
8.3	Reproducibility . . . . .	23
8.4	Contributions . . . . .	23
8.5	Implications . . . . .	23
8.5.1	For Design . . . . .	23
8.5.2	For Society . . . . .	23
8.6	Limitations . . . . .	23
8.7	Future Directions . . . . .	23
8.8	Closing Remarks . . . . .	23
	<b>References</b>	<b>24</b>

<b>Appendices</b>	<b>25</b>
<b>A First appendix</b>	<b>26</b>

## 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 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

As researchers, the selection of appropriate methods is crucial in order that our interpretations may be based in truth. The selection of methods necessarily entails research of its own, with approaches each having distinct advantages and disadvantages, and thus the decisions that were made are discussed below. It is worth noting that different methods may produce different results, meaning this thesis cannot claim to be an entire representation of knowledge, but rather a reflection of the researcher, the ways the research was done, and the particular philosophical environment that fostered the work.

### 3.2 Experimental Methods

In order to understand how changing visual and design features can affect interpretations of scatter-plots, the work presented here tests hypotheses using controlled experiments. This allows for the isolation of factors of interest from other features, and allows us to establish causality with regards to effects. The use of empirical studies in visualisation has a rich history [1], and facilitates the move towards designing from the ground up based on theory, as opposed to a more traditional paradigm in which visualisations are only tested after they have been designed by technical experts.

#### 3.2.1 Experimental Design

Chapter 1

### **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 Experiments**

#### **4.4.1 Experiment 1: Uniform Opacity Adjustments**

**Introduction**

**Methods**

**Analysis**

**Discussion**

#### **4.4.2 Experiment 2: Spatially-Dependent Opacity Adjustments**

**Introduction**

**Methods**

**Analysis**

**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**

## **Scatterplots Optimised for Correlation**

## **Perception Affect Beliefs About Correlations and Attitudes Towards Trivial Topics**

### **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 Experiment: I don't know what is going to happen yet**

#### **7.4.1 Introduction**

#### **7.4.2 Methods**

#### **7.4.3 Analysis**

#### **7.4.4 Discussion**

### **7.5 General 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**



# References

- [1] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. “Empirical Studies in Information Visualization: Seven Scenarios”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.9 (Sept. 2012), pp. 1520–1536. ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.279. (Visited on 03/14/2024) (cited on p. 16).

# Appendices

# **Appendix A**

## **First appendix**

### **A.1 Section in Appendix**