VisiStat: Visualization-driven, Interactive Statistical Analysis

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Krishna Subramanian

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Dr. Pierre Dragicevic

Registration date: 12.02.2014
Submission date: 10.03.2014

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

$Aachen, March 2014$
Krishna Subramanian

Contents

Abstract **xvii**

Acknowledgements **xix**

Conventions **xxi**

1 Introduction **1**

1.1 Statistical Analysis in HCI research 1

1.2 Potential Reasons for this Problem 3

1.3 Contributions of this Thesis 4

2 Related work **7**

2.1 Problems Inherent with Statistical Analysis . 7

2.2 Interactive Visualizations 10

2.3 Standard Statistical Analysis Softwares . . . 12

2.4 StatWing 14

2.5 ViSta 15

2.6 Statistical Graphics 16

Histogram & Bar Chart 19
Dot Plot 20
Stem-and-leaf Display 21
Boxplot 22
Letter-value summary 22
Scatterplot & Scatterplot Matrix . . . 24
Star Plot & Glyph Plot 26
Interaction Plot 27
Biplot 28
Heat Map 29
Pie Chart 29
Sparkline 30
Ternary Chart 31

3 Design Approach **33**

3.1 Iterative Design of VisiStat 33
3.2 Design Principles 35
3.2.1 Minimize prior knowledge needed by the user 35
3.2.2 Make visualizations interactive 35
3.2.3 Disclose information only when required 36
3.3 Pre-design Decisions 36

3.3.1 Data Visualization Plotting Approach 36

Variable-driven approach 37

Visualization-driven approach 37

3.3.2 Choice of Statistical Analysis Tasks . . 40

3.4 Structure of our Preliminary User Study . . . 40

3.4.1 Sessions 41

Performing significance tests 42

Performing transformation to normality 42

Finding correlation between two variables and fitting a regression model for prediction 43

Testing the variable-driven visualization selection approach . . . 45

3.4.2 Participants 45

3.5 Issues and Resulting Changes 46

3.5.1 Selection of Means 46

Old Design 46

Change(s) Made 48

3.5.2 Help 49

Old Design 49

Change(s) Made 52

3.5.3 Scale Showing the Difference in Means 54

Old Design 54

Change(s) Made 56

3.5.4 Visualization for Normality of Distributions 56

Old Design 56

Change(s) Made 57

4 Interaction Design 59

4.1 Preprocessing and Dataset Selection 59

4.2 Exploring Data 60

4.3 Interacting with Visualizations 63

4.4 Automatic Selection of Statistical Tests 64

4.5 Performing Assumption Tests 65

4.6 Viewing the Results 65

4.6.1 Difference in Means 66

4.6.2 Confidence Interval of the Difference in Means 66

4.6.3 Effect Size 68

4.6.4 Other Statistics 68

4.7 Further Analysis 68

4.8 Correlation and Simple Linear Regression . . 70

4.9 Multiple Regression 71

4.10 Interaction Effect 73

4.11 Data Transformations 74

5 Evaluation **77**

5.1 Research Questions 77

5.2 Structure of Final User Study 79

5.2.1 Questionnaire 79

5.2.2 Report Analysis 81

5.2.3 Qualitative Feedback 82

5.3 Participants 82

5.4 Results 83

5.5 Limitations 87

5.6 Discussion 89

6 Summary and future work **91**

6.1 Summary 91

6.2 Future work 93

6.2.1 Implementation 93

6.2.2 User study 94

A Implementation **95**

A.1 Libraries used 95

B User Study **101**

Bibliography **105**

Index **107**

List of Figures

1.1 A Complete Experimental Research 2

1.2 Importance of Visualizations in Statistical Analysis 4

2.1 TouchViz 11

2.2 Result of Statistical Analysis in R 12

2.3 Selection of Statistical Test in SPSS 13

2.4 Boxplots show Data-specific Knowledge . . . 14

2.5 StatWing 15

2.6 Guide maps in ViSta 15

2.7 Visualizations in ViSta. 16

2.8 Results of Statistical Analysis in ViSta () . . . 17

2.9 Histogram 19

2.10 Bar Chart 20

2.11 Cleveland Dot Plot 20

2.12 Wilkinson Dot Plot 21

2.13 Stem-and-leaf Display 22

2.14 Boxplot 23

2.15 Letter-value Summary 23

2.16 Scatterplot 24

2.17 Scatterplot Matrix 25

2.18 Starplot 26

2.19 Interaction Plot 27

2.20 Biplot 28

2.21 Heat Map 29

2.22 Pie Chart 30

2.23 Sparklines 31

2.24 Ternary Chart 31

3.1 Iterative Design of VisiStat 34

3.2 testsInVisiStat 41

3.3 Preliminary User Study: Significance Tests . 42

3.4 Preliminary User Study: Selection Of 2 Means 43

3.5 Preliminary User Study: Transformations to Normality 43

3.6 Preliminary User Study: Correlation 44

3.7 Preliminary User Study: Regression 44

3.8 Preliminary User Study: Variable-driven Visualization Selection 45

3.9 Old Design: Information Scent for Selecting Means 46

3.10 Old Design: Performing Significance Tests . . 47

3.11 Old Design: Issue with Mean Selection 48

3.12 New Design: Mean Selection 50

3.13 Old Design: Tooltips for Accessing Help . . . 51

3.14 New Design: Dedicated Help Button for Accessing Help 51

3.15 New Design: Help Mode 53

3.16 Old Design: Scale Showing Difference in Means 54

3.17 New Design: Scale Showing Difference in Means 55

3.18 Old Design: QQPlot for Visualizing Distributions 56

3.19 New Design: Modified Histograms for Visualizing Distributions 57

4.1 Preprocessing and Dataset Selection 60

4.2 Main Page of VisiStat 61

4.3 Boxplot in VisiStat 62

4.4 Histogram in VisiStat 62

4.5 Interacting with Visualizations in VisiStat . . 63

4.6 Performing Significance Tests in VisiStat . . . 64

4.7 Assumptions in VisiStat 65

4.8 Results of Significance Test in VisiStat 66

4.9 Results: Confidence Interval 67

4.10 Results: Effect Size 67

4.11 Pairwise Post-hoc Test in VisiStat 69

4.12 Tukey's HSD Test in VisiStat 69

4.13 Correlation in VisiStat 70

4.14 Simple Linear Regression in VisiStat 71

4.15 Scatterplot Matrix in VisiStat 72

4.16 Multiple Regression in VisiStat 73

4.17 Interaction Plot in VisiStat 74

4.18 Data Transformation in VisiStat 75

5.1 Structure of Final User Study 80

5.2 Comparison of Scores in Report Analysis Tasks 85

B.1 Questionnaire to assess Users' Level of Expertise on Statistical Analysis Tasks 102

List of Tables

2.1 Issues found in reported statistics (Cairns [2007]) 9

2.2 The prevalence of visualizations was established by surveying prominent literature on statistical analysis. 18

3.1 The visualizations in VisiStat and their utilities 38

3.2 Selection of visualizations from number and type of variables 39

3.3 Solutions that were considered for the issue: selection of means 49

3.4 Solutions that were considered for issue: scale showing the difference in means 55

5.1 Table showing the options for each question in the questionnaire and its score 81

5.2 Number of errors is balanced between pre-test and post-test report analysis tasks. 82

5.3 The coding guidelines for rating annotations collected from the report analysis tasks . . . 88

A.1 List of Statistical Tasks 98

A.2 List of statistical assumptions and the tests used to validate them. 99

B.1 Sample Annotations and Expert Ratings . . . 101

B.2 Types of errors in the Report Analysis task with examples 103

B.3 Inferred behavior of participants in the final user study 104

Abstract

In various disciplines, statistical analysis plays a vital role as it lends evidence to claims made by the author. Over the years, several studies have indicated a lack of adequate statistical knowledge among researchers. Further analysis reveals problems, some of which are inherent with statistical analysis. As an attempt to solve these problems, we introduce VisiStat: visualization-driven, interaction statistical analysis tool. VisiStat lessens the complexities of statistical analysis by embedding knowledge upfront. Users are guarded against making mistakes with statistical analysis by collocation of appropriate visualizations that act as precursor to statistical analysis. VisiStat also enables users to interact with visualizations and thereby improves user's ability to develop interests.

In our user study, we found VisiStat to be an effective statistical analysis tool, which helps users solve 90% of their research questions. VisiStat also empowers users to perform statistical analysis tasks that they have not done before. In addition, VisiStat shows promise to be used as a learning tool, which can improve users' statistical knowledge. In this Thesis, we explain the development of VisiStat in detail and summarize our findings.

Acknowledgements

I would firstly like to thank Chat Wacharamanotham, M.Sc., for without his guidance this work would not have been possible. I am immensely grateful to him for his support and advice during the course of this Thesis.

I would like to thank the wonderful people at the Media Computing Group, for giving me valuable feedback and for readily participating in my user studies.

I would like to thank James Eagan, for his insightful comments that opened up new dimensions for my Thesis.

I would like to thank Prof. Dr. Jan Borchers, my thesis advisor, and Dr. Pierre Dragicevic, my second examiner, for their time and support.

Finally, I would like to thank my parents, sister, and friends for their moral support during the course of this Thesis.

Thank you!

Krishna Subramanian.

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in coloured boxes.

> **EXCURSUS:**
> Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Definition: *Excursus*

The whole thesis is written in American English.

Chapter 1

Introduction

"Statistical Thinking will one day be as necessary for efficient citizenship as the ability to read and write."

- H. G. Wells.

In this chapter, we discuss the concern of inadequate statistical knowledge among HCI researchers and its consequences on HCI research and researchers alike. We then discuss the potential causes of this problem and show how VisiStat can help assuage this problem.

1.1 Statistical Analysis in HCI research

Statistical analysis is an important part of HCI research

Statistical analysis plays a crucial role in HCI research (Fig. 1.1). A complete experimental research in HCI is like a three-legged stool that has these legs:

1. A valid research hypothesis
2. A well-designed experiment that collects data pertinent to research hypothesis
3. Clear, unambiguous statistical analysis of the experimental data that validates the research hypothesis

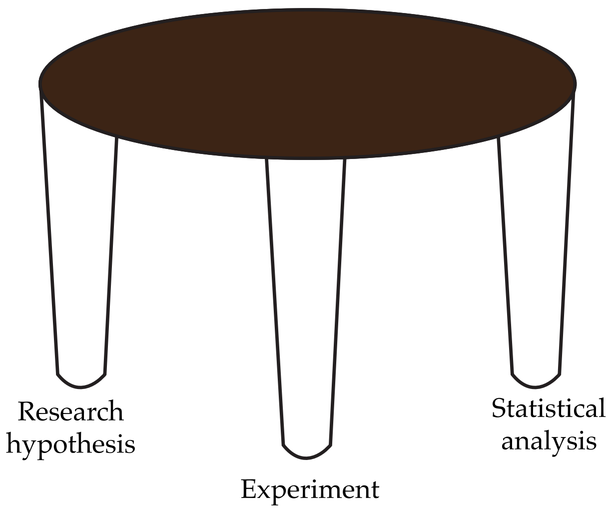


Figure 1.1: A complete experimental research is analogous to a three legged stool with research hypothesis, experiment, and statistical analysis as its legs.

Hypothesis, experiment, and statistical analysis are vital elements of experimental research in HCI

Like the three-legged stool, the research breaks down when one or more of these supporting legs are stunted or broken. E.g., research without a valid research hypothesis does not address a concern of interest and may not have a reliable set of related work it is built upon; research without a well-designed experiment does not assess the research hypothesis in a meaningful manner; and finally, research without a proper statistical analysis does not have a valid claim.

Inadequacy of quality statistical analysis in HCI

Over the years, several studies and reports have indicated an insufficient quality of statistical analysis in HCI research. E.g., a survey conducted on 41 research papers, sampled from leading HCI journals, found that 40 of them had one or more issues with statistics (Cairns [2007]). Some of the common issues found were inappropriate testing, disregard to statistical assumptions, over testing, and issues with reporting.

The claims made by the authors of such issue-ridden research papers are subject to debate. Some issues can be serious enough to falsify the findings. Further, this also refutes the claims made by authors, who use the issue-ridden research papers as a credible resource to base their findings on. Given how important the role of statistics in HCI research is, this is a pressing issue and one that must be dealt with sooner than later.

1.2 Potential Reasons for this Problem

Several reasons can be attributed to this problem:

- Lack of formal statistical education,
- Vastness of statistical analysis, and
- Misgivings of statistical analysis.

Researchers, who do not have a strong mathematical background, do not usually undergo formal training in statistical analysis. They resort to other methods of learning such as reading books, attending seminars, workshops, online courses, and so on. However, these are not quick fixes. E.g., an online course, *Data Analysis and Statistical Inference*[1], which teaches the basics of statistical analysis, has a course duration of 10 weeks. This is not desirable for a researcher, who has a limited amount of time at her disposal.

Learning statistical analysis takes a lot of time

Statistical analysis is a vast, intricate field and demands a large amount of prior knowledge from users. To perform statistical analysis, a typical user needs to have knowledge of:

Users need vast amount of prior knowledge as prerequisite

1. Selecting the appropriate statistical test based on the experimental design, data types of the variables, and roles of the variables.
2. The list of statistical assumptions for the chosen test and the tests to validate them. Additionally, the user needs to know the alternatives to be used if one or more of the assumptions are violated.

[1]https://www.coursera.org/course/statistics

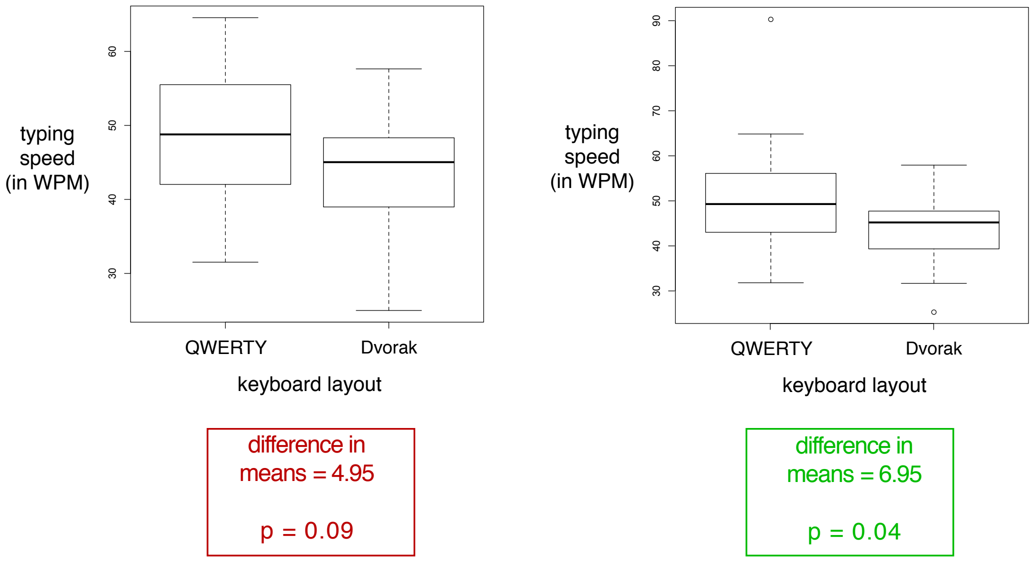


Figure 1.2: This image shows why statistical analysis, in isolation, can be misleading. The presence of an outlier in the box plot on the right, makes the effect significant!

3. An understanding of the various statistical analysis terms such as p-value, effect size, and test statistic. This is required to interpret the results of statistical analysis test.

1.3 Contributions of this Thesis

This need for a vast amount of prior knowledge is a huge stumbling block for beginners. To reduce the amount of prior information required by them, we embed the required knowledge with VisiStat. We will elaborate on this knowledge in Section 2.3 "Standard Statistical Analysis Softwares".

Statistical analysis needs visualizations to complement them

Statistical analysis can be misleading when considered in isolation (Fig. 1.2). Data-specific knowledge such as presence of outliers, shape of the distributions, and variance of the distributions, is also needed to gain a complete understanding of the data. For this purpose, VisiStat collocates the results of statistical analysis with appropriate visualizations. This makes the user be aware of data-specific knowledge and reduces the mistakes by improving understanding. Details of some of the common statistical analysis mistakes are discussed in Section 2.1 "Problems Inherent with Statistical Analysis".

Further, we introduce a novel approach of performing statistical analysis by interacting with visualizations. By initiating statistical analysis using visualizations, users initially explore the data to find potential anomalies such as incorrect data-entry, mistakes with data transformation, and so on, with it. Also, previous research has shown that interactivity promotes users' involvement in exploring data and improves users' skill to develop insights (Rzeszotarski and Kittur [2013], Perer and Shneiderman [2008]).

Statistical analysis is initiated by interacting with visualizations

To summarize, VisiStat has the following contributions to the field of HCI:

1. It removes the need for prior knowledge from the users by embedding the knowledge required by them upfront in the system.
2. It improves users' understanding of data and helps avoid mistakes by collocating results of statistical analysis with appropriate visualizations.
3. By allowing users to interact with visualizations to perform statistical analysis, VisiStat promotes a healthy practice of Exploratory Data Analysis and increases their ability to develop insights.

Chapter 2

Related work

In the previous chapter, we discussed the motivation behind VisiStat by elaborating some reasons behind the lack of adequate statistical analysis knowledge among HCI researchers. In this chapter, we take a closer look at some problems with statistical analysis and show how VisiStat tries to solve them. We will then discuss research done in the fields of interactive visualizations, statistical analysis softwares, and statistical graphics.

2.1 Problems Inherent with Statistical Analysis

VisiStat was inspired by a study that investigated the quality of reported statistics in HCI (Cairns [2007]). The study aimed to find issues with statistics that were reported in research papers. 81 research papers from leading journals: Human-Computer Interaction (HCIJ '06), British Computer Society Human-Computer Interaction (BCS HCI '05, '06), and ACM Transactions on Computer Human Interaction (TOCHI '06) were examined in this study. Of these, roughly half of them (41 research papers) reported statistics. Note that, this indicates the prevalence of statistics in HCI.

A study that found there were issues with reported statistics

The study found that 40 (98%) research papers had issues with the reported statistics. Incidentally, the one paper that did not have any issues reported only simple descriptive statistics involving correlation. The issues found in the research papers are listed in Table 2.1 "Issues found in reported statistics (Cairns [2007])".

The seriousness of this finding is subject to the importance of the finding, type of issue (e.g., reporting error is considerably less critical than over testing), and influence of the research paper (the journal in which it is published, number of citations it has, and so on). However, the survey does indicate that the quality of statistical analysis in HCI research is low.

In addition to this study, other research works hint at the same issue:

Some common issues with statistical analysis and counter-measures

Zuur et al. [2010] discuss some of the common issues with statistical analysis such as the detrimental effect of outliers, shape and spread of distributions, interaction effect, and so on. The authors also suggest measures to counter these issues. E.g., to detect outliers, an appropriate visualization such as box plot or Cleveland dot plot can be used.

In VisiStat, users can use box plot to detect outliers. In addition, VisiStat also employs histograms to inform the user about the shape and spread of distributions, and uses interaction plots to indicate possible interaction between independent variables.

NHST is prone to a lot of issues

Johnson [1999] discusses some of the inherent problems with NHST (Null Hypothesis Significance Testing) such as issues with assumptions of NHST, misinterpretation of statistical significance as practical significance, arbitrariness of p-value, and so on. The author also proposes alternative methodologies such as use of confidence intervals and effect sizes, modeling, Bayesian approach, decision theory and so on.

In VisiStat, effect sizes and confidence intervals are reported in addition to p-value (Section 4.6.3 "Effect Size"). Since, p-value is displayed at the bottom of the results, and is not visualized, we aim to motivate the user to rely more

Issue	Example	Comments
Reporting	*Users typed significantly faster with QWERTY keyboard layout when compared to Dvorak keyboard layout (p <0.001)*	The test that was used, mean and standard deviation of the distributions that were compared are missing.
Checking Assumptions	*Paired t-test revealed that users typed significantly faster with QWERTY keyboard (M = 30 WPM) layout when compared to Dvorak keyboard layout (M = 20 WPM) (p <0.001)*	There is no indication that assumptions were satisfied (normality of distributions and homogeneity of variances).
Over-testing	*Paired t-test revealed that users typed significantly faster with QWERTY keyboard (M = 30 WPM) layout when compared to Dvorak keyboard layout (M = 20 WPM) (p <0.001)*	Over-testing without adjusting the significance level, increases the chances that "some chance variation will be extreme enough to seem like a significance." (Cairns [2007])
	One-way ANOVA revealed that users typed significantly faster with QWERTY keyboard (M = 30 WPM) layout when compared to Dvorak (M = 20 WPM) and i10 (M = 22 WPM) keyboard layout (p <0.001)	
Inappropriate testing	*Paired t-test revealed that users typed significantly faster with QWERTY keyboard (M = 30 WPM) layout when compared to Dvorak keyboard layout (M = 20 WPM) (p <0.001)*	As the experimental design does not change for keyboard layout against typing speed, it is incorrect to use paired t-test for one case and an unpaired t-test for another case.
	Unpaired t-test revealed that users typed significantly faster with i10 keyboard (M = 30 WPM) layout when compared to Dvorak keyboard layout (M = 20 WPM) (p <0.001)	

Table 2.1: Issues found in reported statistics (Cairns [2007])

on effect sizes and confidence intervals to interpret results.

Students do not understand statistical analysis concepts

Garfield and Ahlgren [1988] discuss concern over unawareness of statistical analysis concepts among students, even at college level. They trace the reasons for this problem to inadequate mathematical knowledge, abstract reasoning, and, partly, to the nature of statistical analysis itself. The authors recommend several ways to improve the students' learning experience: use of visual illustrations and simulations over abstractions (e.g., tabular output), warn students of common statistical analysis mistakes and guard them against it, and so on.

VisiStat uses several visual illustrations to communicate statistical analysis concepts to the user. E.g., error bars for confidence intervals, interpretable bar for effect sizes, visualizations for assumptions, and so on.

We can see from this analysis that existing research points towards the existing design and features of VisiStat.

2.2 Interactive Visualizations

TouchViz uses physical metaphors to allow user interaction with data

TouchViz is a multi-touch visualization tool that uses physical metaphors such as gravity, momentum, and force in its design to allow users to interact with data (Fig. 2.1). To evaluate this tool, the authors conducted a user study, in which they initially educated their participants on various tools that are available in TouchViz (clustering, filtering, histograms, and scatterplots). The participants were then asked to identify as many patterns in a given dataset as possible. A post survey was conducted at the session to gather feedback.

Interactivity promotes exploration and fun

Participants reported TouchViz to be visually appealing and fun to use. Authors also noticed that the participants exhibited "ludic behavior", which improves users' performance and cognition. This indicates that interactivity furthers Exploratory Data Analysis.

TouchViz focuses primarily on visualizations and the scope

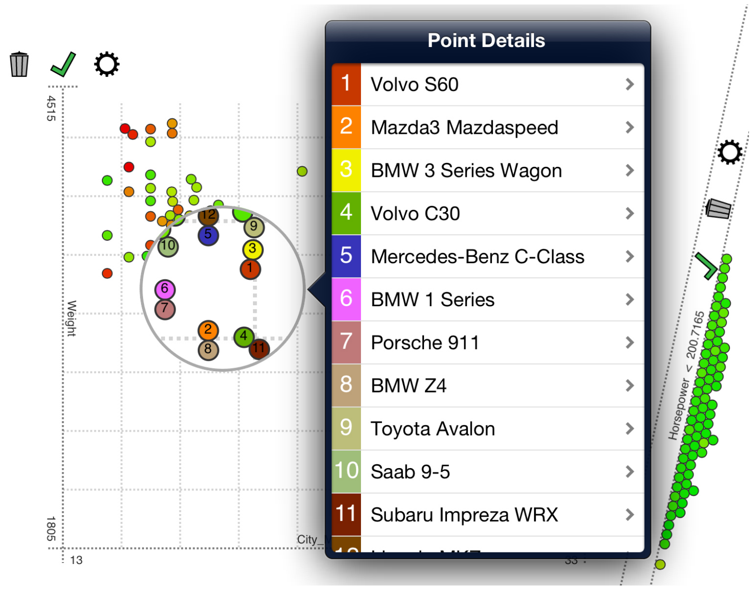


Figure 2.1: *Lens* or *razor* can be used to filter data points in TouchViz (Rzeszotarski and Kittur [2013])

of statistical analysis is limited to elementary tasks such as filtering and clustering.

Like TouchViz, VisiStat is also interactive, albeit in the domain of web-based Graphical User Interface, where the interaction is limited to use of mouse and keyboard.

Other tools and APIs that support interactivity

Visualization tools such as Tableau and Many Eyes are widely used across the world to interactively explore data. However, the interaction is limited to revealing descriptive statistics such as means, medians, standard deviations, and so on. Libraries like InfoVis, prefuse, D3, and VisPy enable developers to develop interactive visualization softwares. However, unlike VisiStat, these focus primarily on visualizations and descriptive statistics.

```
> model = lm(typingSpeed ~ keyboardLayout, kbd)
> anova(model)
Analysis of Variance Table

Response: typingSpeed
               Df Sum Sq Mean Sq F value    Pr(>F)
keyboardLayout  2 1193.1  596.57  11.816 5.092e-05 ***
Residuals      57 2877.8   50.49
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure 2.2: Knowledge of statistical analysis concepts is needed to interpret results of statistical analysis in R.

2.3 Standard Statistical Analysis Softwares

Users need a lot of knowledge to work with existing softwares

Existing softwares like R, SPSS, and JMP act as the de facto standard for statistical analysis in HCI. These are powerful tools that enable users to perform a multitude of statistical analysis tasks. However, these require the user to have a large amount of prior knowledge:

- Selection of statistical tasks from experimental design and variables' roles (Fig. 2.3). E.g, if the user wants to find the statistical test for comparing the effect of a between-group factor of categorical, independent variable with 2 levels on a ratio, dependent variable, she must know that an Unpaired t-test should be used.
- List of statistical assumptions for a statistical test and the tests to validate them. Additionally, the list of alternative tests that can be used when one or more assumptions are violated. E.g., the user must know that an Unpaired t-test requires normality of distributions, homogeneity of variances, and independence of samples to be satisfied. The user should also know that Mann-Whitney U test and Welch's t-test can be used as alternatives for Unpaired t-test.
- Interpretation of statistical analysis terms that is needed to interpret the results of statistical analysis

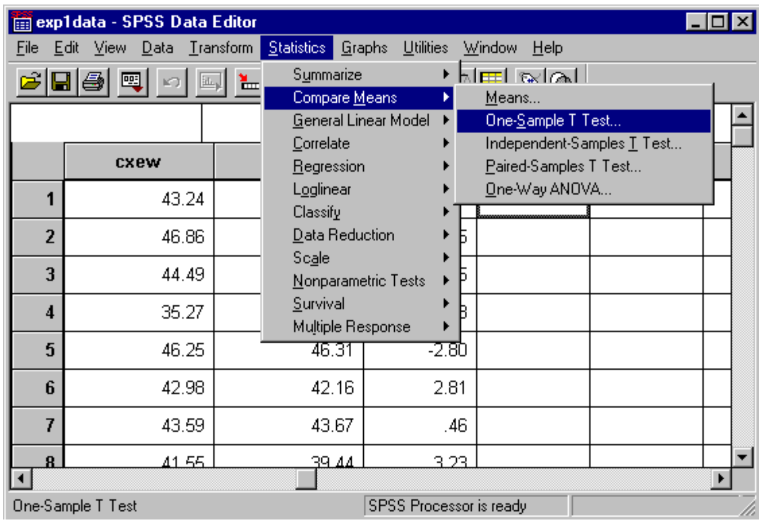


Figure 2.3: Selecting appropriate test in SPSS requires knowledge of experimental design and variables' roles.

(Fig. 2.2). E.g., to interpret the results of an Unpaired t-test, the user should know what Cohen's d, p-value, and students t-statistic are.

To acquire this knowledge, users need spend a lot of time on tutorials or books to use such statistical analysis softwares, which is not desirable.

These statistical analysis softwares offer tools to visualize data. However, these visualizations are not used to initiate statistical analysis. Also, they are often isolated from statistical analysis. E.g., to perform an Unpaired t-test in R, the user does not need to view any visualization and can initiate the statistical test directly. The user is potentially unaware to anomalies in the data such as outliers, data-entry and measurement errors, and so on. E.g., a box plot in Fig. 2.4 shows that there is an extreme data point, which could affect the significance test performed.

Visualizations and statistical analysis are separated in existing softwares

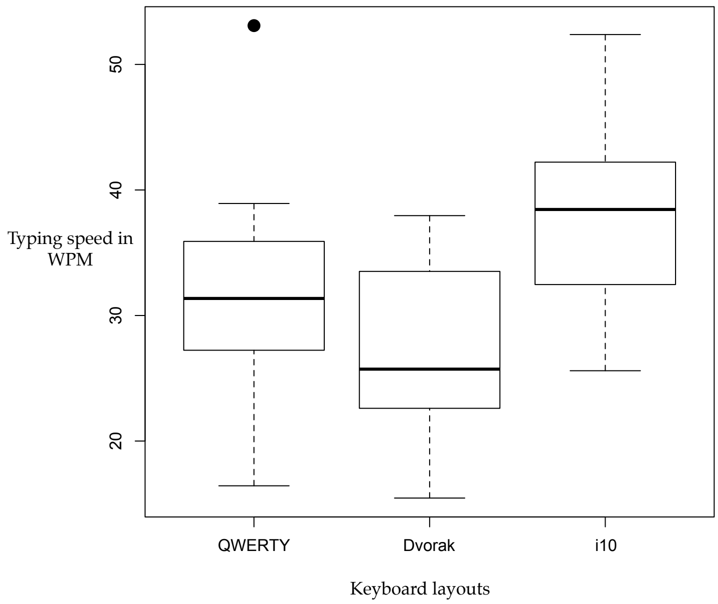


Figure 2.4: Box plots can be used to identify anomalies in the data such as the presence of outliers.

2.4 StatWing

StatWing uses knowledge in the world to simplify statistical analysis

StatWing is a web-based, commercial software that vastly simplifies statistical analysis by automatically selecting the appropriate statistical analysis task based on the selected variables. E.g,. when the user selects a categorical variable and a numerical variable, StatWing infers that the user wants to do a t-test (Fig. 2.5).

Unlike VisiStat, StatWing does not allow users to interact with the visualizations to perform statistical analysis. As a result, users are not exposed to data-specific knowledge that could help them avoid potential misinterpretations.

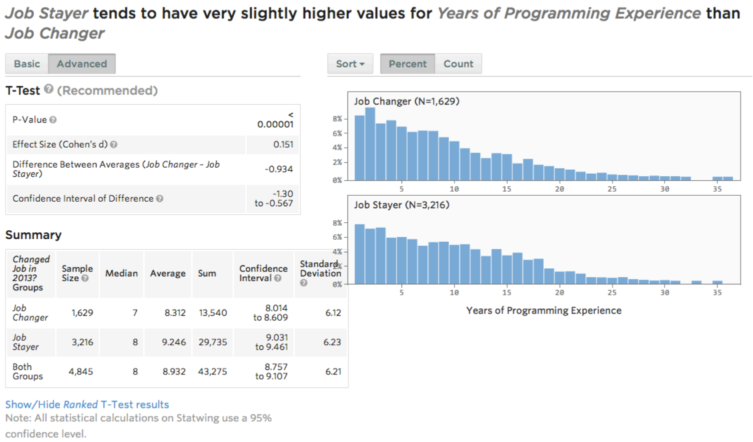


Figure 2.5: StatWing infers the appropriate statistical analysis task from the selected variables.

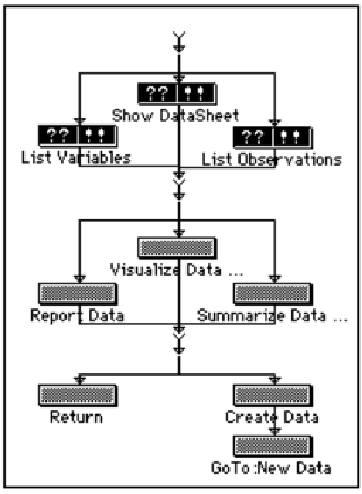


Figure 2.6: Guidemaps are used to guide the user through different steps of statistical analysis (Young and Bann [1996]).

2.5 ViSta

ViSta is an open-source tool that allows user to perform statistical analysis (Young and Bann [1996]). It uses visual graphics, called *guidemaps*, to guide the users to perform

ViSta uses visualizations to guide users to perform analysis

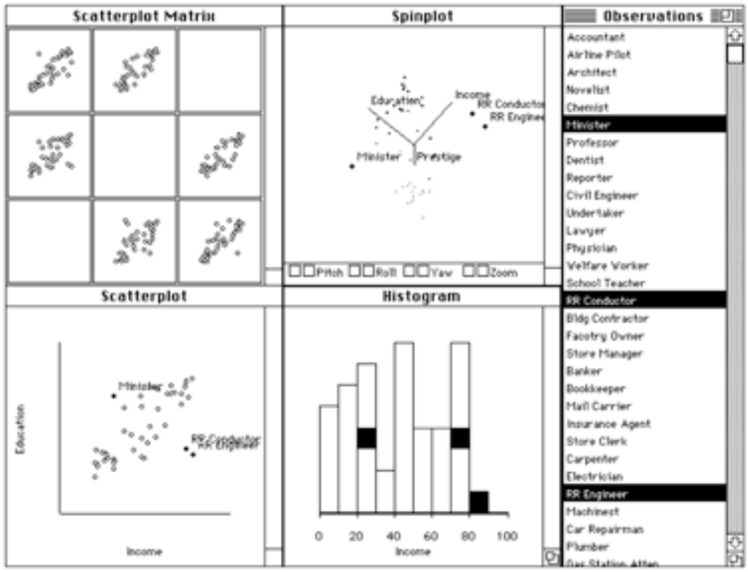


Figure 2.7: Visualizations in ViSta do not lead to further analysis (Young and Bann [1996]).

statistical analysis. The *guidemap* shows the sequential list of tasks that are required to do analysis. E.g., *show dataset, list variables, visualize data, and summarize data* are some of the tasks (Fig. 2.6).

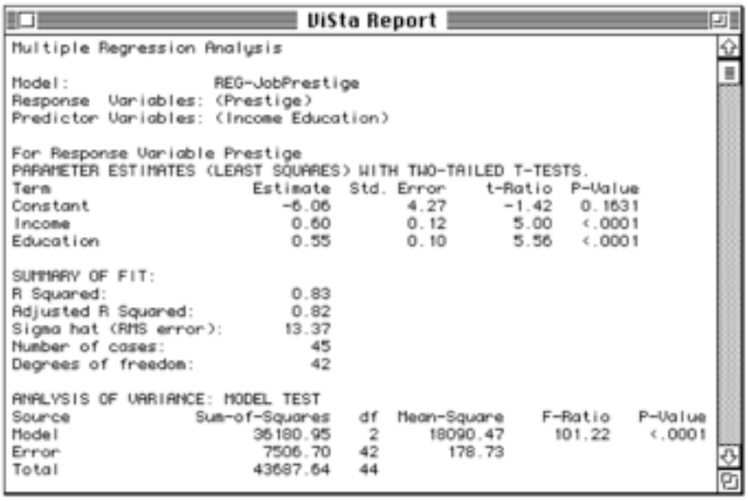
The data visualizations are optional and do not lead to further analysis (Fig. 2.7). Like existing statistical analysis softwares, statistical analysis serve as separate entities that are isolated from appropriate visualizations (Fig. 2.8).

2.6 Statistical Graphics

Statistical graphics or visualizations are needed to completely understand the data

Statistical graphics deal with the use of visualizations that helps the user to analyze data. They *"let us see what may be happening over and above what we have already described"* (Tufte and Graves-Morris [1983]). The integration of statistical methods and visualizations improves the ability of the user to develop insights (Perer and Shneiderman [2008]).

For use in VisiStat, we surveyed a variety of statistical



UiSta Report

Multiple Regression Analysis

Model: REG-JobPrestige
Response Variables: (Prestige)
Predictor Variables: (Income Education)

For Response Variable Prestige
PARAMETER ESTIMATES (LEAST SQUARES) WITH TWO-TAILED T-TESTS.

Term	Estimate	Std. Error	t-Ratio	P-Value
Constant	-6.06	4.27	-1.42	0.1631
Income	0.60	0.12	5.00	<.0001
Education	0.55	0.10	5.56	<.0001

SUMMARY OF FIT:

R Squared:	0.83
Adjusted R Squared:	0.82
Sigma hat (RMS error):	13.37
Number of cases:	45
Degrees of freedom:	42

ANALYSIS OF VARIANCE: MODEL TEST

Source	Sum-of-Squares	df	Mean-Square	F-Ratio	P-Value
Model	36180.95	2	18090.47	101.22	<.0001
Error	7506.70	42	178.73		
Total	43687.64	44			

Figure 2.8: Results of statistical analysis are isolated from visualizations (Young and Bann [1996]).

graphics. These graphics were collected from various sources: Behrens [1997], and existing softwares such as Tableau, R, and SPSS. The following criteria were then used to select the visualizations:

- As the visualizations were used to initiate statistical analysis, visualizations should complement statistical analysis tasks (Fig. 3.2). E.g., to perform a significance test, the appropriate visualization is one that shows the central tendency (mean) and the variance of the distribution.
- Visualizations should help the user gain data-specific knowledge (e.g., presence of outliers, shape of distribution, variance of distribution, and interaction effect).
- As the target users of VisiStat are beginners, the visualizations should be prevalent and easy to interpret. To determine if a visualization was prevalent, we surveyed books on statistical analysis (Griffiths [2008], Crawley [2005], Field [2009]). E.g., histogram is prevalent, whereas a Cleveland dot plot is not as prevalent (Table 2.2 "The prevalence of visualizations

was established by surveying prominent literature on statistical analysis.”).

- While it was important to provide the user with an arsenal of powerful visualizations to explore data, we also wanted to reduce *redundancy* to prevent users from experiencing “*paradox of choice*” (Schwartz [2004]). For this purpose, visualizations should be functionally non-repetitive from other chosen visualizations.

Visualization	Crawley [2005]	Griffiths [2008]	Field [2009]
Bar chart	✓	✓	✓
Boxplot	✓	✓	✓
Histogram	✓	✓	✓
Scatterplot	✓	✓	✓
Line chart		✓	✓
Pie chart		✓	
Interaction Plot	✓		✓
QQ-plot			✓
Dot plot		✓	
Stem-and-leaf display		✓	
Density plot			✓
PP-plot			✓
Letter-value summary			
Star plot			
Biplot			
Heat map			
Sparkline			
Ternary chart			

Table 2.2: The prevalence of visualizations was established by surveying prominent literature on statistical analysis.

We will now briefly discuss the statistical graphics that

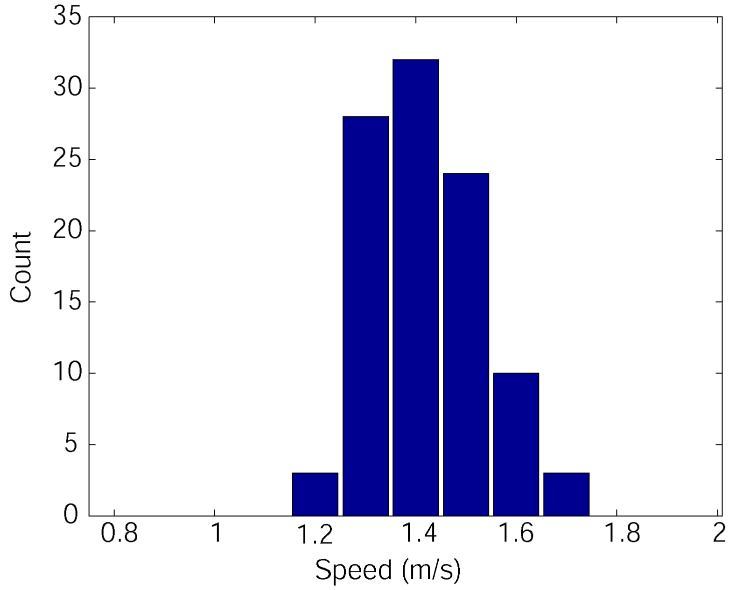


Figure 2.9: Histogram visualizes the distribution of numerical data.

were surveyed for use in VisiStat and the reasons why there were selected or rejected.

Histogram & Bar Chart

Histograms are used to visualize the data distributions (Fig. 2.9). Specifically, they show the shape, skew, and kurtosis of distributions and can also be used to determine the number of modes in a given distribution. To show the frequency distribution of categorical data, bar charts can be used (Fig. 2.10).

Histograms are used to visualize distribution

✓ Histograms and bar charts do not directly complement any statistical analysis tasks. However, they provide data-specific knowledge such as shape of distribution and outliers.

✓ Histograms and bar charts are commonly used in many visualization softwares.

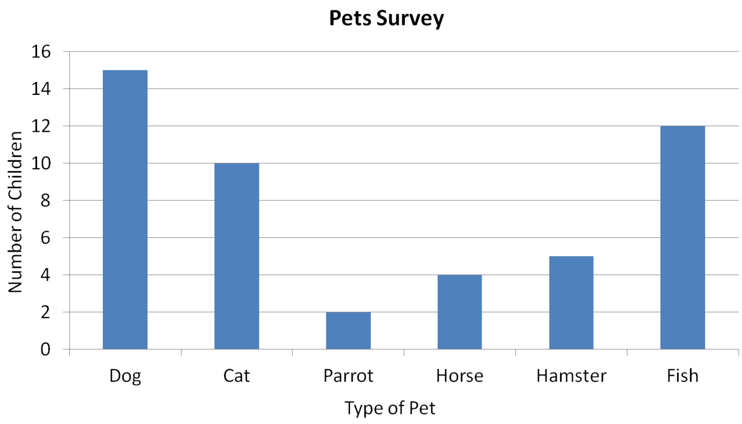


Figure 2.10: Bar chart visualizes the distribution of categorical data.

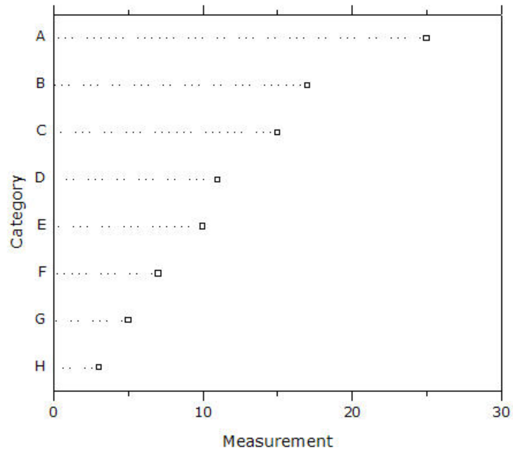


Figure 2.11: Cleveland dot plot can be used as an alternative to bar chart

Dot Plot

Dot plots are used to find outliers

There are two different versions of dot plots: Wilkinson dot plot (Fig. 2.12), which is used for numerical data and Cleveland dot plot (Fig. 2.11), which is used for categorical data. Wilkinson dot plots differs from histogram in that the data points are not distributed in a uniform manner along the horizontal axis. Cleveland dot plots can be used as an alter-

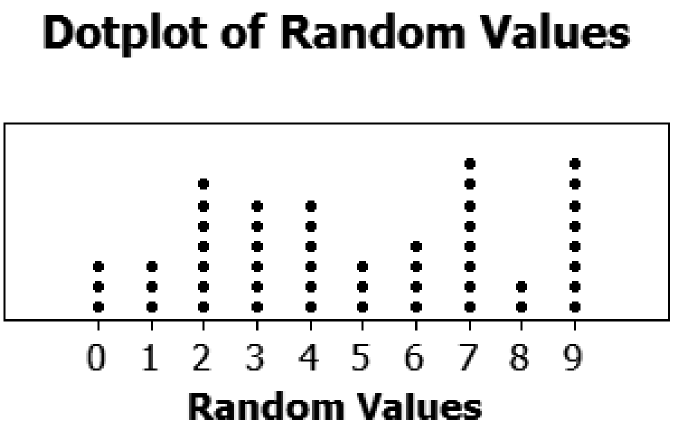


Figure 2.12: Wilkinson dot plot can be used to visualize the distribution of numerical data

native for bar charts and it reduces clutter when compared to bar charts.

- ✗ Although dot plots help in providing data-specific knowledge, they do not offer significant benefits over histograms.
- ✗ Neither versions of dot plot are commonly used in visualization softwares.

Stem-and-leaf Display

Stem-and-leaf displays are also used to visualize distributions (Fig. 2.13). They are similar to histograms in that they cluster the given data points into bins.

Stem-and-leaf displays are used to visualize distribution

- ✗ Although stem-and-leaf displays help in showing data-specific knowledge, they do not offer significant benefits over histograms.

0	2
1	1 2 3
2	3
3	1
4	
5	2
6	5
7	
8	
9	4
10	
11	
12	3

Figure 2.13: Stem-and-leaf display shows the distribution of data. In this figure, we see that the distribution is skewed towards lower digits (0-3)

✗ Stem-and-leaf displays are not commonly used in visualization softwares.

Boxplot

Boxplots are used to visualize distribution

Boxplots are used to visualize distributions (Fig. 2.14). However, they differ from histograms in that they are better at showing the central tendency, spread, and outliers in distribution.

✓ Boxplots can be used to initiate significance test, which requires information about central tendency (mean) and standard deviation.

✓ Boxplots are commonly used in existing softwares.

Letter-value summary

Letter-value summaries are used to visualize distribution

Letter-value summaries are conceptually similar to boxplots in that they divide the distribution into several fractions (quartiles, eighths, sixteenths, and so on) (Fig. 2.15).

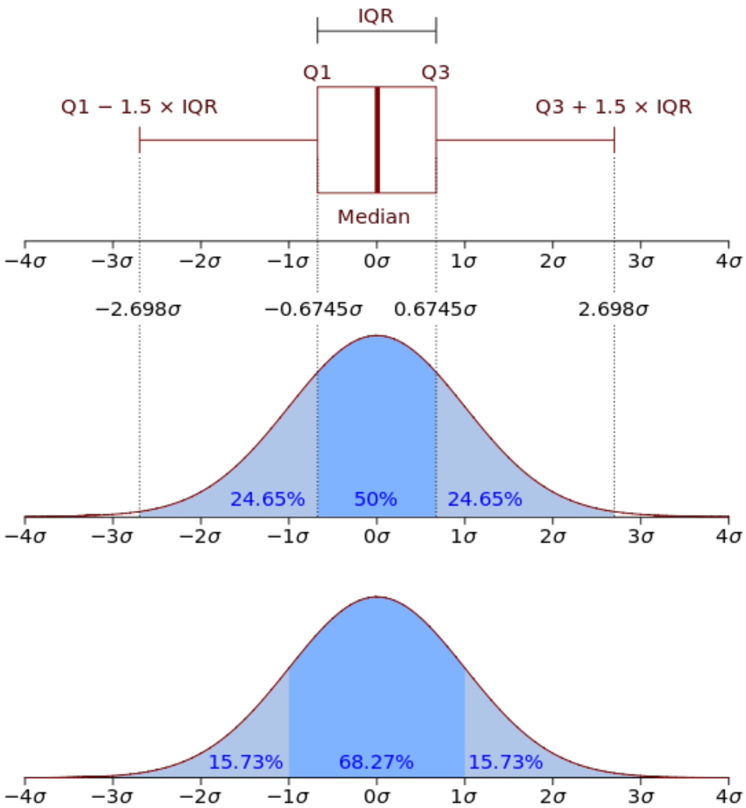


Figure 2.14: Boxplots are used to interpret the spread or variance of a given distribution. In this image, it is compared against Probability Distribution Function of a Gaussian distribution for better understanding.

LETRVALS IMR

	DEPTH		LOWER	UPPER		SPREAD	MIDPT
M	51	-	60.6	60.6	-		60.6
H	26	-	26.2	129.4	-	103.2	77.8
E	13H	-	16.9	166.3	-	149.4	91.6
D	7	-	12.8	200	-	187.2	106.4
1	1	-	9.6	650	-	640.4	329.8

Figure 2.15: Letter-summary is used for interpreting distribution by splitting the data points into several fractions each assigned to a letter

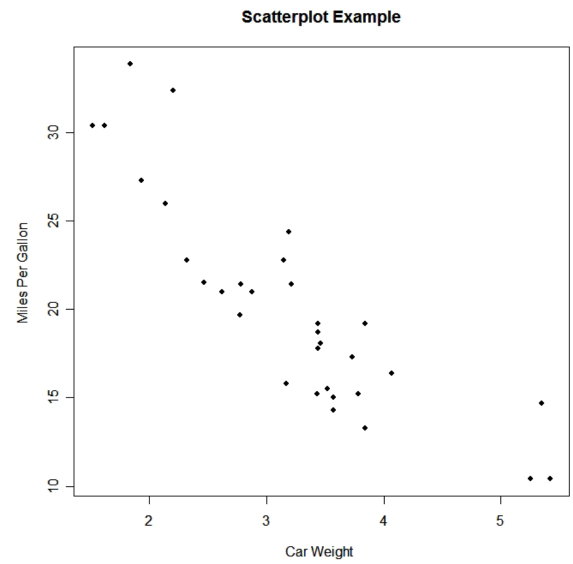


Figure 2.16: A scatterplot shows the relationship between two variables. This figure shows that the variables are negatively correlated.

The fractions are represented by different letters and the data points are assigned to corresponding letters.

✗ Letter-value summaries are not easy to understand for beginners.

✗ They not commonly used in existing softwares.

✗ They do not offer significant benefits over boxplots.

Scatterplot & Scatterplot Matrix

Scatterplots are used to view relationship between variables

A scatterplot is used to view the relationship between two variables (Fig. 2.16). Data is visualized as points on a XY-graph with the coordinates given by values of the two variables. Scatterplots can be extended to include a third variable (usually an independent variable or a dependent variable with finite set of unique values), by using colors (color

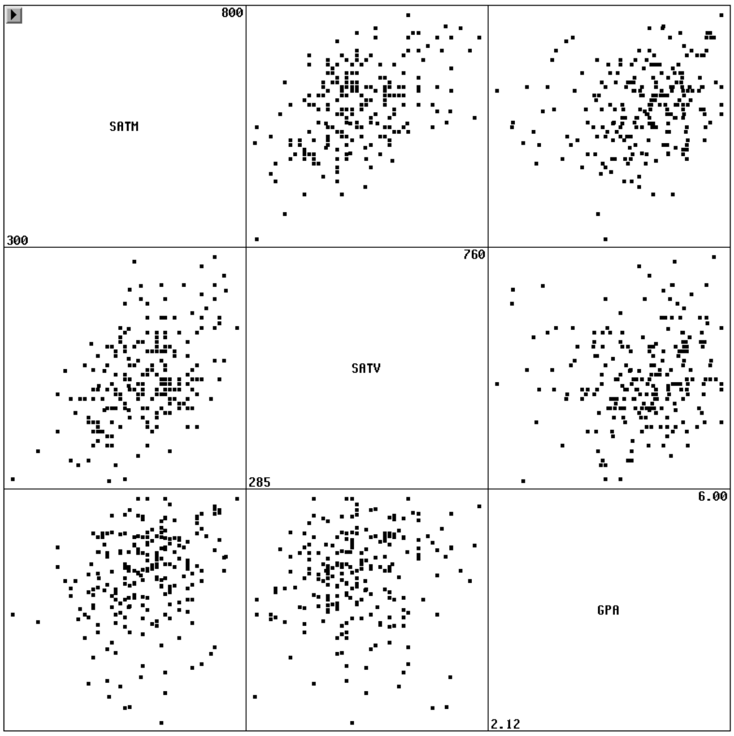


Figure 2.17: A scatterplot matrix is an extension of scatterplot that shows the pairwise relationship between multiple variables.

plot) or radii (bubble plot) to represent the different levels of the third variable.

A scatterplot matrix is an extension of scatterplot to multiple variables (Fig. 2.17). It is a square matrix of order n, where n is the number of variables. Each cell $C(i, j)$, where $i \neq j$, shows the scatterplot between variable i and variable j.

Scatterplot matrix is an extension of scatterplot to multiple variables

✓ Scatterplots complement the statistical analysis task of correlation as they show the relationship between two variables.

✓ Scatterplots are commonly used and are also easy to interpret.

✓ Scatterplot matrices are used to quickly view corre-

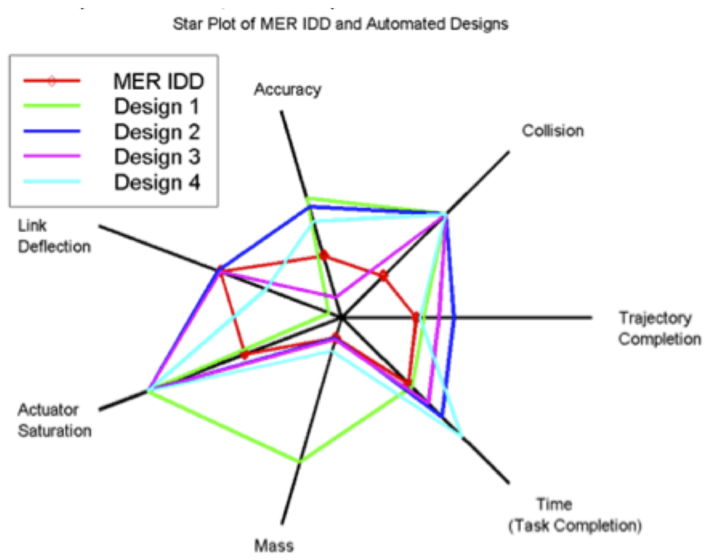


Figure 2.18: A star plot is used to visualize multiple data values for a set of participants.

lations of multiple variables. They can also be used to initiate multiple regression, which is based on relation between multiple variables.

Star Plot & Glyph Plot

Star plots are used to plot multiple variables

Star plots are used to visualize compare three or more variables simultaneously (Fig. 2.18). Spokes or radii are used to represent variables and subjects or participants are represented by colored lines. Star plots can also used to identify outliers. However, they are not good choice of visualization for comparing different participants. They are also not scalable.

- ✗ Star plots do not complement any of VisiStat's statistical analysis task.
- ✓ It can be used to show data-specific knowledge (outliers).
- ✗ Star plots are not prevalent.

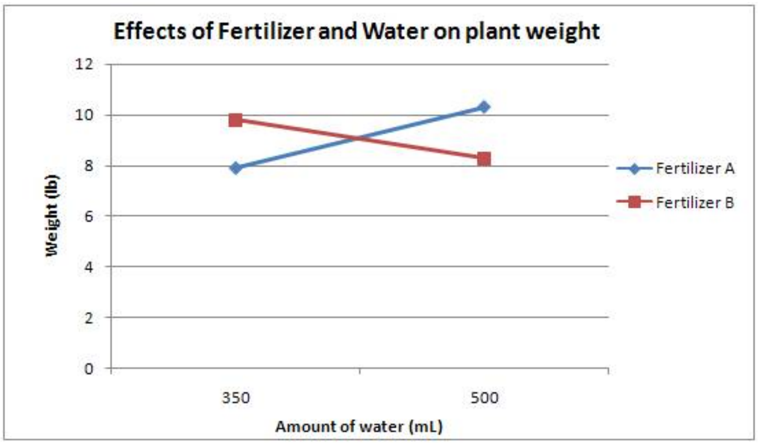


Figure 2.19: Interaction plot shows that there is an interaction between amount of water and plant weight

We discarded star plot as it was not versatile enough - it requires multiple variables and is not scalable to large number of participants. Glyph plot, which is a matrix of star plots, was also discarded.

Interaction Plot

An interaction plot is used to check for the presence of an interaction effect, which is said to have occurred when the effect of an independent variable on a dependent variable differs for different levels of a second independent variable (Fig. 2.19). Interaction plots are discussed in Section 4.10 "Interaction Effect".

Interaction plot is used to check for interaction effect

- ✓ Interaction plot is used to visualize interaction effect, which is a critical part of significance test with multiple independent variables.
- ✓ It shows data-specific knowledge (presence of an interaction), which might help user detect anomalies with data.
- ✗ Interaction plot is not commonly used in visualization tools. It is, however, part of statistical analysis

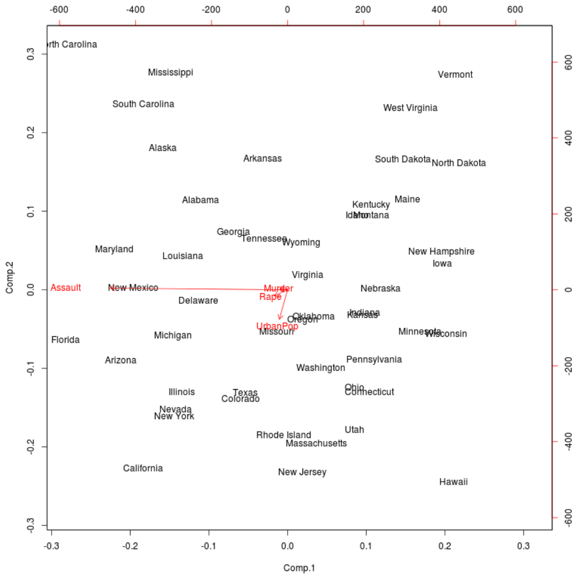


Figure 2.20: A biplot is used to identify outliers and clusters.

softwares.

In VisiStat, interaction plot is used only when multiple independent variables are chosen and when user wants to check the data for the existence of an interaction effect.

Biplot

Biplots are used for cluster analysis and finding outliers

A biplot is a generalized scatterplot that can be used to visualize large number of variables (Fig. 2.20). It displays the data points as well as the variables. Constructing biplots is complex and involves Singular Value Decomposition. Biplots can be used to identify outliers and clusters.

✗ Biplots are used for cluster analysis, which is out of the scope of VisiStat.

✓ It shows data-specific knowledge (outliers) and is better than scatterplot for high-dimensional data.

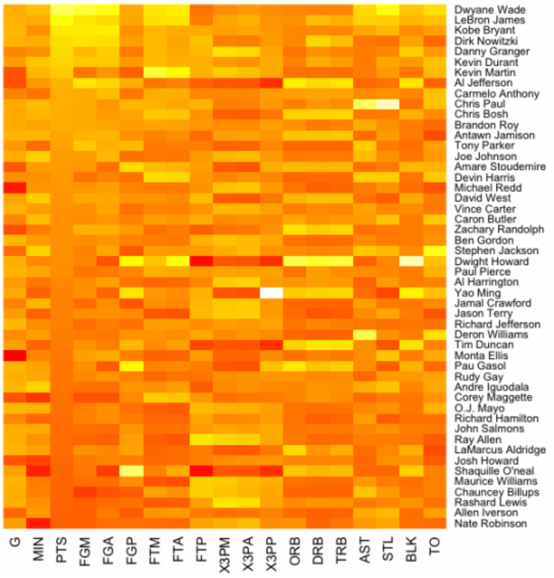


Figure 2.21: Heat map is used to detect patterns and clusters in data

✗ Biplot is not easy to interpret and is not commonly used in visualization softwares.

Heat Map

Heat map is used to visualize a data matrix (Fig. 2.21). It is used to find clusters or patterns with the given data.

Heat maps are used to visualize data matrix

✗ Heat map is used for cluster analysis, which is out of the scope of VisiStat.

✗ It is not versatile and is useful only for analyzing data matrices.

Pie Chart

Pie chart is used to visualize proportions of data (Fig. 2.22). The area or the arc length of a sector (or slice) represents the

Pie charts are used to visualize proportions

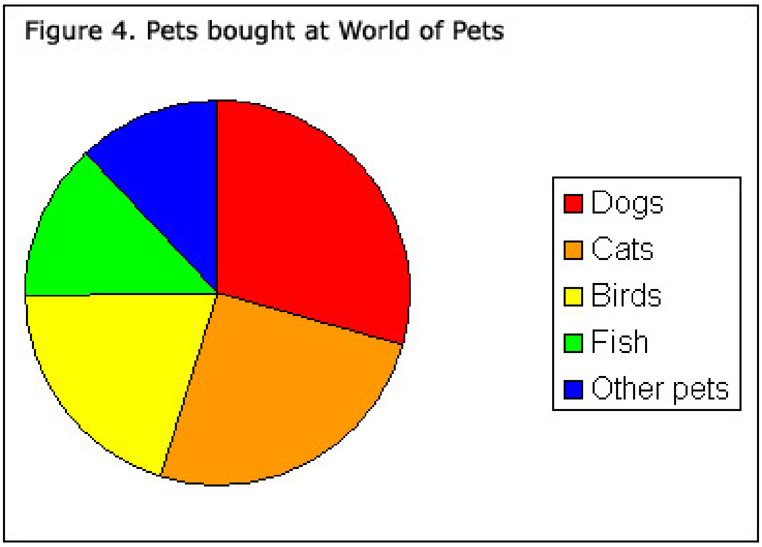


Figure 2.22: Pie chart visualizes the ratios or proportions of data

proportion of the corresponding variable or level of variable.

- ✓ Pie charts are commonly used in visualization softwares.
- ✗ Comparing different sectors visually is difficult and prone to errors.

Sparkline

Sparklines are line graphs for time series analysis

Sparkline is a simple line graph that is typically used for time series analysis (Fig. 2.23). It is used to visualize variations in given data.

- ✗ Sparklines do not offer significant benefits over histograms.

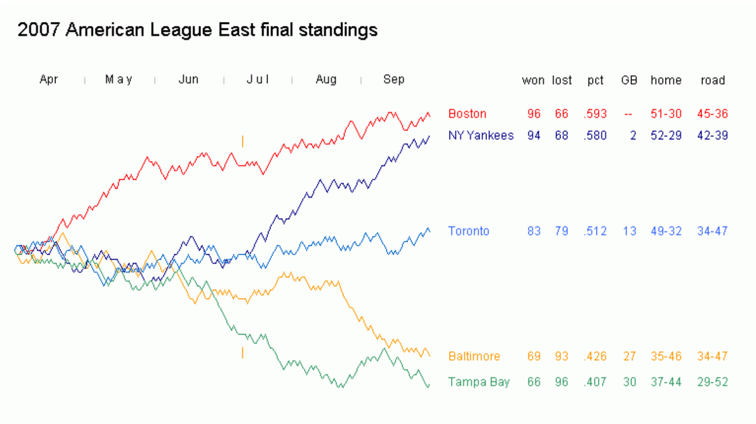


Figure 2.23: Sparkline shows variation in data. It is typically used for time series data.

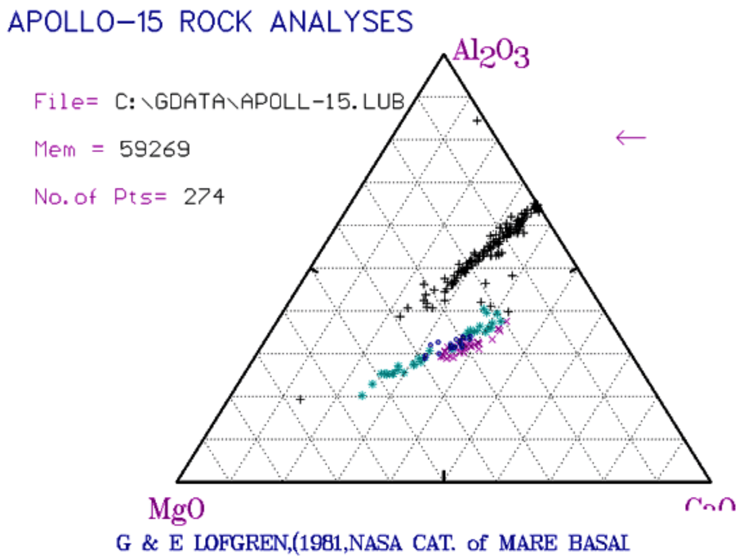


Figure 2.24: Ternary chart shows ratio of 3 variables on an equilateral triangle.

Ternary Chart

Ternary chart is used to visualize ratios of 3 variables on an equilateral triangle (Fig. 2.24). It is used in Petrology, Population Genetics, Metallurgy, and similar fields.

Ternary chart is used for visualizing proportions

✗ It can only be used when comparing ratios of three variables and, therefore, is not versatile.

In addition to these graphics, we discarded the following:

- Visualizations that are used for time-series analysis
  - *Run chart* and *seasonal sub-series plot*.
  - *Recurrence plot* and *Poincaré plot*.
- Visualizations that serve a specific purpose
  - *Control chart*, *p-chart*, and *np-charts* are used to analyze process control.
  - *Forrest plot* and *Galbraith plot* are used to compare treatment effects across multiple studies.
  - *Funnel plot* is used to identify publication bias in research.

From our analysis, the following visualizations were finally chosen to be part of VisiStat:

1. Histogram
2. Boxplot
3. Scatterplot
4. Scatterplot matrix
5. Interaction plot (as a part of statistical analysis and not for initial exploration)

Chapter 3

Design Approach

In the previous chapter, we discussed existing research work done in the fields of interactive visualization, statistical analysis, statistical graphics, and some inherent issues with statistical analysis. In this chapter, we discuss the evolution of VisiStat's design over three iterations of design, prototyping, and evaluation.

3.1 Iterative Design of VisiStat

Since VisiStat uses a novel interaction design, we used DIA cycle-of-development to develop it incrementally (Fig. 3.1).

VisiStat follows an iterative design

DIA CYCLE-OF-DEVELOPMENT:
In a Design-Implement-Analyze cycle, a prototype is designed and a implemented. This prototype is evaluated and changes are then incorporated into the design of next iteration. The process continues for several iterations till the end product is developed.

Definition:
DIA cycle-of-development

At the early stages of development, we developed a low-level paper prototype. This focused on the interaction design for performing statistical tests: significance tests, correlation, and regression. It was evaluated by using expert feedback.

Low-level prototype was evaluated by expert feedback

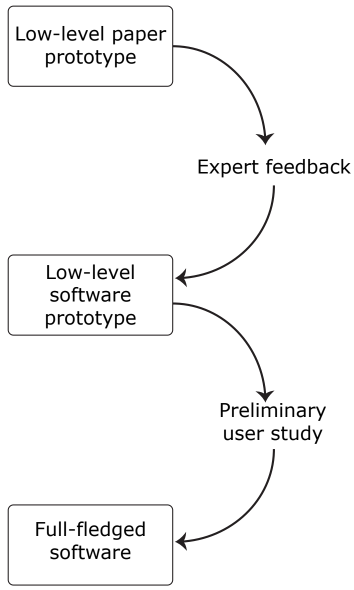


Figure 3.1: The evolution of VisiStat over iterations of design, prototyping, and evaluation.

A medium-level, interactive, mock-up prototype was then designed based on the feedback. It consisted of separate interfaces for performing the following tasks:

- Selecting means to perform a significance test
- Performing transformation to normality and then performing a parametric significance test
- Finding correlation between two variables and fitting a regression model for prediction
- Testing the variable-driven visualization selection approach

Medium-level prototype was evaluated by a controlled experiment where users followed a think-aloud protocol

To evaluate this prototype, we conducted a user study that aimed to find potential issues with interaction design. In our user study, participants were required to perform a certain set of tasks with each interface (Refer Section 3.4.1 "Sessions"). We will now discuss the design principles involved in VisiStat followed by some pre-design decisions that were made. We will then discuss our preliminary user study and the issues identified from the subsequent analysis. Finally, we will discuss the changes that were made to VisiStat to address these issues.

3.2 Design Principles

VisiStat is built around three core design principles: *minimize prior knowledge needed by the user*, *make visualizations interactive*, and *disclose details only when needed*. We will now briefly discuss these principles.

3.2.1 Minimize prior knowledge needed by the user

VisiStat aims to minimize the amount of prior knowledge that is needed by the user to perform statistical analysis. As we discussed in Section 2.3 "Standard Statistical Analysis Softwares", users require a vast amount of knowledge to perform and interpret statistical analysis. VisiStat reduces this burden by embedding the required knowledge upfront. E.g., selection of data visualizations and statistical analysis tasks is automated based on experimental design, variables' types, and variables' roles (Fig. 3.2). This enables novice users, who do not have a lot of statistical knowledge, to perform statistical analysis. VisiStat also tests the statistical assumptions of the chosen statistical test automatically, and performs the appropriate test accordingly.

By embedding knowledge in the system, VisiStat minimizes prior knowledge needed by the user

3.2.2 Make visualizations interactive

Unlike existing softwares, in VisiStat, users interact with visualizations to perform statistical analysis. This way, users are made aware of potential anomalies with their data such as the presence of outliers, data-entry mistakes, measurement errors, and so on.

Visualizations allow users to identify anomalies with data

Also, interactivity develops the user's ability to develop insights, which helps the user to perform Exploratory Data Analysis. From Rzeszotarski and Kittur [2013], interactivity also promotes user's involvement in analyzing data.

3.2.3 Disclose information only when required

Definition: *Progressive disclosure*

PROGRESSIVE DISCLOSURE:
The concept where hidden content is shown "*just in time*", when and where user need it. (Tidwell [2010])

Detailed information is revealed only when needed

In progressive disclosure, the features of the system are split into primary and secondary sets. The primary set of features are used frequently by the user and are therefore readily accessible. The secondary set of features, however, are not often needed and are delegated to the next stage in interaction. The features of the secondary set can be accessed only on-demand.

E.g., in VisiStat, when the tests of statistical assumptions are performed, only an indication of whether assumptions have been satisfied or violated is initially revealed. Only when a violation occurs or upon user's request, additional details of the test such as the visualization are disclosed.

3.3 Pre-design Decisions

Prior to the implementation of VisiStat, we made a few design decisions:

3.3.1 Data Visualization Plotting Approach

Unlike many existing statistical analysis softwares, visualizations are at the heart of VisiStat. They are initially used to analyze the data to develop insights and identify potential anomalies. Subsequently, they used to initiate statistical analysis by interaction.

There are two approaches for plotting visualizations. We discuss these approaches and present our rationale for choosing variable-driven visualization approach.

Plotting a visualization requires the following information:

- Type of visualization
- Variables needed for the visualization

Based on the order in which the information is specified, visualizations can be plotted in two ways:

Variable-driven approach

In a variable-driven approach, the variables are specified first. Based on the number and type of variables that are specified, the visualization is chosen (Table 3.2 "Selection of visualizations from number and type of variables"). In VisiStat, we infer the statistical task that the user wants to perform from the selected variables and then select appropriate visualization to complement the statistical task.

Visualizations are chosen automatically based on variables

✓ It minimizes user's need to have knowledge of visualizations and their utilities (Table 3.1 "The visualizations in VisiStat and their utilities"). The user only needs to select the variables that she is interested in exploring.

Visualization-driven approach

In this approach, the type of visualization is specified and the variables are then chosen to plot the visualization. Most existing softwares use this approach.

Variables are fitted to chosen visualizations

✗ User needs to be aware of visualizations and their utilities.

Visualization	Purpose
Histogram	• View the shape, skew, and kurtosis of a distribution • Check the modality of a distribution • Identify outliers
Boxplot	• View the spread or variance of distribution • View the central tendency of distribution (mean, median) • Identify outliers
Scatterplot and Scatterplot matrix	• View the pairwise relationship between variables (correlation)

Table 3.1: The visualizations in VisiStat and their utilities

VisiStat uses a variable-driven visualization selection approach due to the following reasons:

- This approach does not require users to know about visualizations and their utilities. Novice users often do not have this knowledge.
- When users work with statistical analysis softwares, they have the objective of solving certain research questions. Variable-driven approach offers direct mapping from information that user has (research questions, which consists of variables) to user's desired goals (solving research questions). E.g., for the research question "*Do users type significantly faster with the QWERTY keyboard layout when compared to Dvorak layout?*", the user selects the independent vari-

able *keyboardLayout* and the dependent variable *typingSpeed*. From Table 3.2 “Selection of visualizations from number and type of variables”, box plot is selected as the default visualization, which further leads to significance test between the two distributions.

Table 3.2 “Selection of visualizations from number and type of variables” shows the visualizations that are automatically selected based on the number and type of variables selected. Note that, in our implementation of VisiStat, all dependent variables were assumed to be numerical (ratio or interval) and all independent variables were assumed to be categorical (nominal or ordinal).

Number of Independent Variables	Number of Dependent Variables	Selected Visualization	Alternative Visualizations
0	1	Histogram	Boxplot
	2	Scatterplot	Histogram
			Boxplot
	3+	Scatterplot Matrix	Histogram
			Boxplot
1	0	Histogram	-
	1	Boxplot	Histogram
			Scatterplot
	2	Scatterplot (color)	Scatterplot matrix
	3+	Scatterplot Matrix	-
2	0	Scatterplot	Scatterplot Matrix
	1	Boxplot (color)	Scatterplot (color)
			Scatterplot Matrix
	2+	Scatterplot Matrix	-
3+	0+	Scatterplot Matrix	-

Table 3.2: Selection of visualizations from number and type of variables

3.3.2 Choice of Statistical Analysis Tasks

Literature on statistics in HCI was reviewed to identify common tasks

Since implementation of all the existing statistical analysis tests was not feasible, we had to limit the scope of VisiStat's statistical analysis tasks. For this purpose, we surveyed some prominent literature on statistics in HCI (Statistics for HCI Research[1], Lazar et al. [2010], Practical Statistics for HCI[3]) to identify common statistical analysis tasks used in HCI. We further limited the implemented tasks by considering only numerical dependent variables and categorical independent variables. For a complete overview of tests implemented in VisiStat, see Fig. 3.2. Additional tests can be easily integrated to the existing design and this is planned as a future work (Section 6.2.1 "Implementation").

3.4 Structure of our Preliminary User Study

User study to find issues with interaction design; think-aloud protocol; logged audio, screen

This section describes the structure of our preliminary user study. The aim of this study is to identify possible issues with user interaction. For this purpose, we followed a think-aloud protocol and logged the audio, video, and screen recordings. There were no time constraints given to participants. Before the session, participants were given some background information about VisiStat. Also, for each session, the experimenter briefly explained the dataset and visualizations involved in the session.

For each session, the experimenter kept track of the tasks that the participant had done. After session, participants were debriefed about the tasks that they had either done incorrectly or not done. The experimenter then gathered feedback from them, which was later used to identify issues with the interaction and to develop possible solutions for these issues.

[1]http://yatani.jp/HCIstats/HomePage[2]
[3]http://depts.washington.edu/aimgroup/proj/ps4hci/[4]

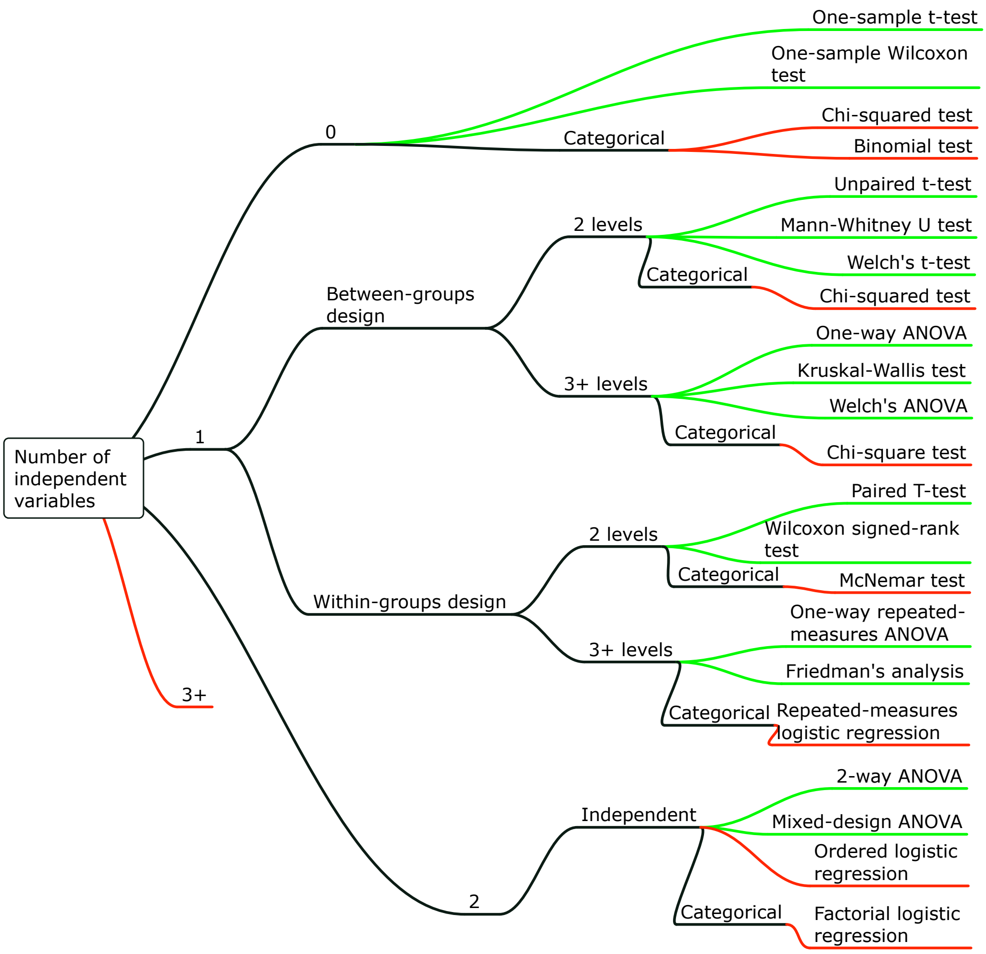


Figure 3.2: Shows the list of tests implemented in VisiStat.

3.4.1 Sessions

We will give an overview of the sessions in our study in the following paragraphs. The details of the interaction are elaborated in the next section, along with the issues and solution.

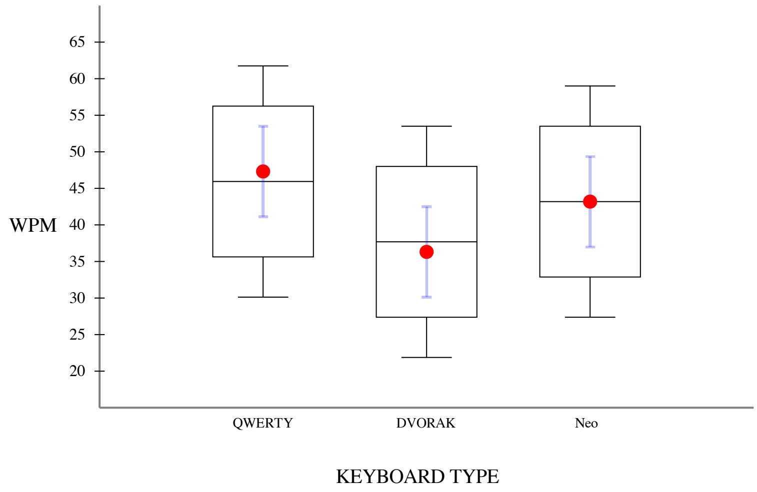


Figure 3.3: Task: perform significance test to compare the given distributions

Performing significance tests

Session to find issues in selection of means

In this session, participants were required to perform a significance test for the three distributions and then, for just two distributions. To perform these, the user had to select the means of distributions by clicking on them (Fig. 3.3). The set of tasks also included interaction with other elements of interface such as confidence intervals, help, mean, and so on. We also noted if participants were able to interpret these elements.

Performing transformation to normality

Session to find issues in transformation

In this session, participants were asked to perform a t-test by selecting the means of the two given means (Fig. 3.4). However, unlike the previous session, one of the distributions was not normal. Participants then were required to transform the distribution to normal distribution (Fig. 3.5). Therefore, to successfully complete this session, participants had to identify that one of the distributions was not normally distributed, and perform transformation to get the results of parametric test.

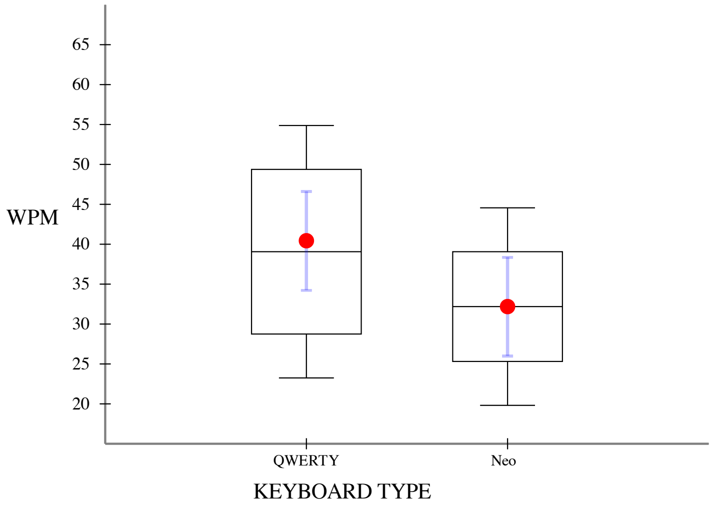


Figure 3.4: Task: perform significance test to compare the given distributions

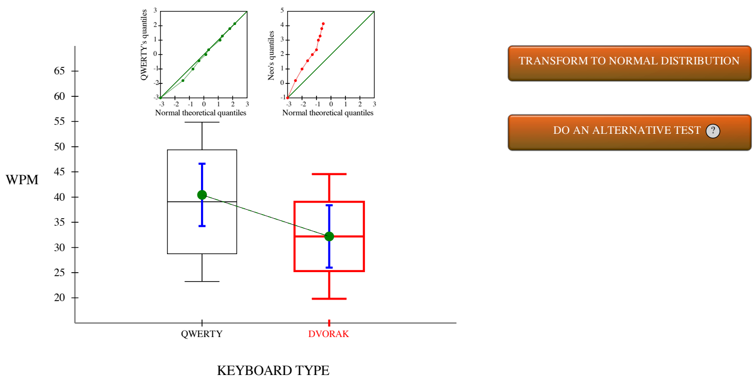


Figure 3.5: Task: transform data to normal distribution

Finding correlation between two variables and fitting a regression model for prediction

In this session, the participants were required to find the correlation between two variables (Fig. 3.6). They were then asked to fit a regression model to predict the value of an outcome variable from an explanatory variable (Fig. 3.7). Like the previous sessions, they were also required to

Session to find issues in correlation and regression

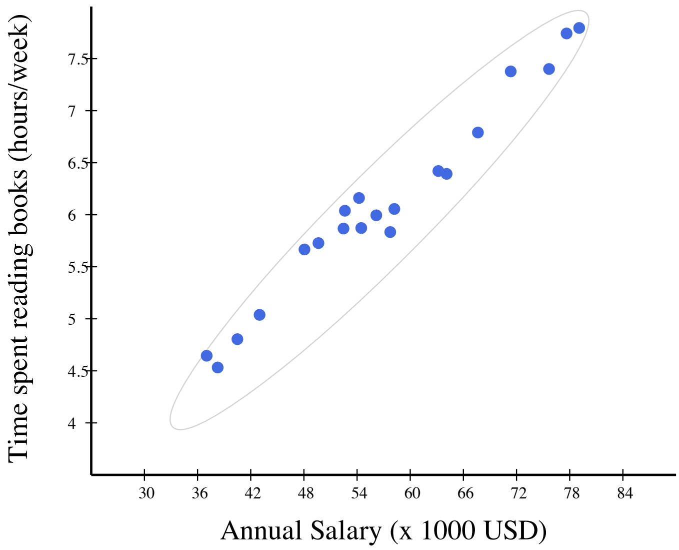


Figure 3.6: Task: find correlation between two variables

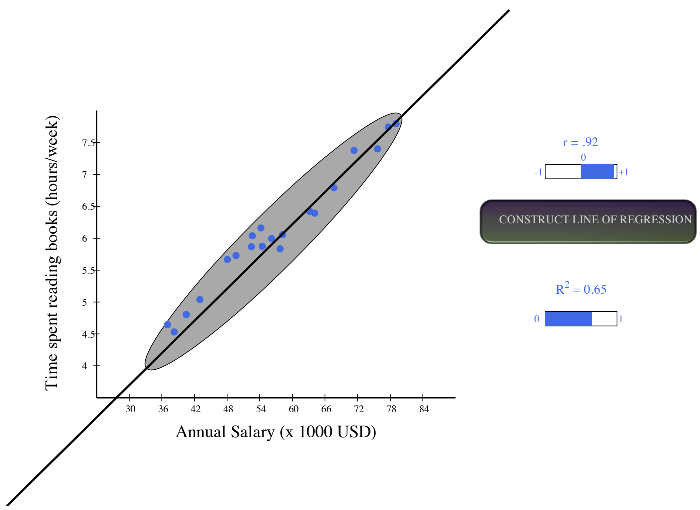


Figure 3.7: Task: predict the value of an outcome variable from explanatory variable.

interpret the concepts involved.

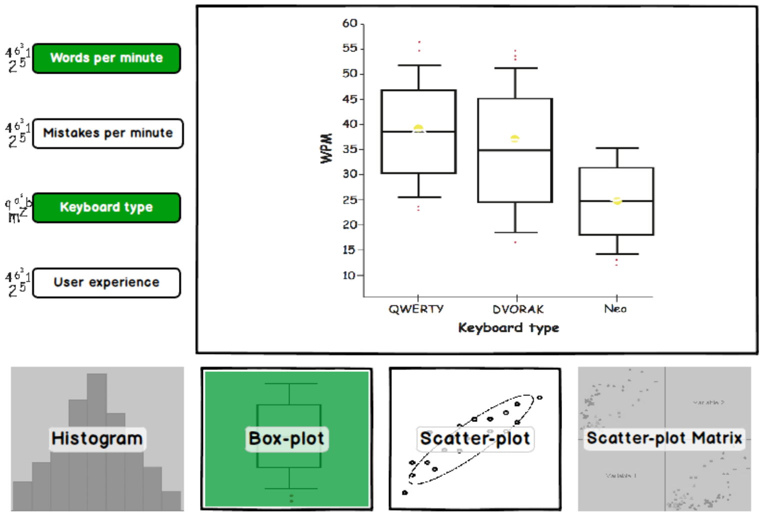


Figure 3.8: Task: plot certain visualizations by selecting specified variables.

Testing the variable-driven visualization selection approach

The purpose of this session is to investigate the variable-driven visualization selection approach. To test this, the participants were given randomized combinations of variables and visualizations as targets and were asked to navigate to them (Fig. 3.8). Any issues with the navigation were noted down and clarified later with the participant.

Session to find issues in variable-driven visualization selection

The entire user study lasted 35 minutes for each participant on average.

3.4.2 Participants

7 users (1 female) were recruited from the local university for this study. Their ages ranged from 24 to 26. All users had some basic knowledge of experimental design. However, not everyone had knowledge of statistical analysis.

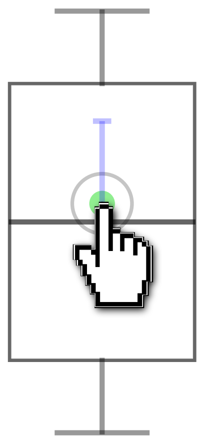


Figure 3.9: Hovering with the mouse pointer over mean induces user to click on the mean.

3.5 Issues and Resulting Changes

In this section, we discuss the issues identified from the study and the changes made to VisiStat based on users' feedback. In order to identify the issues for each task, we analyzed the frequency of users who did a task incorrectly or did not interpret it.

3.5.1 Selection of Means

Old Design

Users were guided by VisiStat to perform significance tests

To perform a significance test, users had to select the means of the distributions they want to compare. VisiStat guided the users to do this by employing *information scents* at various points in the interaction. E.g., an aura of animated, concentric circles induces users to click on means (Fig. 3.9).

There was no distinction between selection phase and comparison

Clicking on a mean selects it for comparison in a signifi-

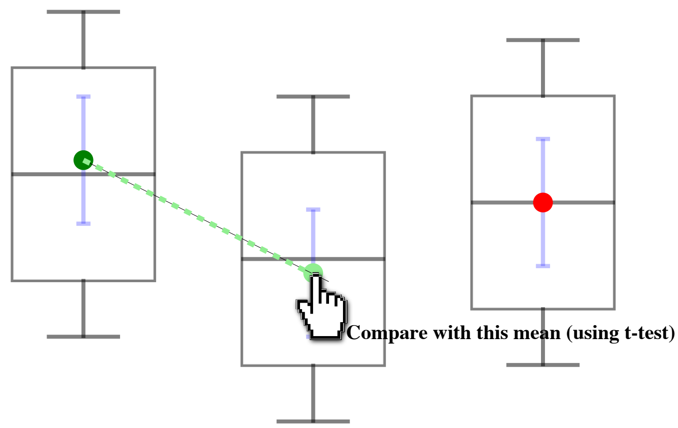


Figure 3.10: Information scents are used to guide user to select means to perform significance test.

cance test. Users can select more distributions in a similar manner, guided by VisiStat's information scents. When the user moves the mouse pointer away from a selected mean, a red line is drawn from the selected mean to the mouse pointer. At this instance, information scent on the other means guides the user to select them. When the user hovers over such means, the red line turns green, inducing the user to select it (Fig. 3.10). At this point, the statistical task that will be performed upon mouse click is also displayed to the user.

Pros & Cons

✗ After selecting two means for comparison, users can either perform a t-test with the selected distributions or continue selecting means. To cancel the *selection* phase and to compare the selected means, users had to click with the mouse pointer outside the means (Fig. 3.11). In our user study, 6 users were unable to do this. Hence, The separation between the *selection* phase and the *comparison* phase is not obvious.

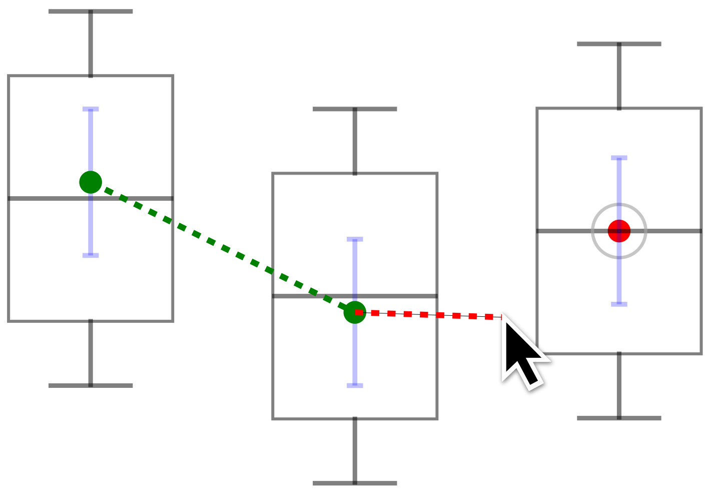


Figure 3.11: Canceling selection phase was not intuitive to 6 users.

Change(s) Made

Users enter a selection mode to select means for comparison

Users can enter a *selection* mode, where they can select or unselect distributions for comparison by left-clicking on the means (Fig. 3.12). After the selection has been made, users can do the comparison by left-clicking on a button.

Comparison with the Old Design

✓ There is a clear separation between the *separation* and the *comparison* phase.

✗ The interaction involves additional steps (e.g., clicking on buttons). However, addition of *selection* buttons has made the selection process more efficient.

Solutions Considered	Pros & Cons
As soon as the users select a mean, the results of the significance test that is appropriate for the current selection can be displayed. On further selection, the results are updated.	✗ The users still had to cancel the *selection* phase.
The users do any of the following to cancel the *selection* phase: • Use a keyboard shortcut (E.g., Esc) • Do a right-click • Double-click on a mean to end selection	✗ These solutions do not present any clear affordances. The users need to *learn* them.

Table 3.3: Solutions that were considered for the issue: selection of means

3.5.2 Help

Old Design

The in-app help in VisiStat, aims to help novice users who do not have much statistical knowledge. In the old design, we tested two approaches of accessing help. In our study, all users were able to discover and then use this approach for accessing help. However, they did so at different points of time in the interaction.

Two approaches to help - tooltips and dedicated buttons - were assessed

1. **Tooltips**

   In this approach, users hover the mouse pointer over elements in the interface to access help (Fig. 3.13). The help text is displayed near the mouse pointer (and, apparently, near the element for which the help was accessed).

   Pros & Cons

   ✓ This approach does not clutter the user interface.

   ✓ As the help text is shown near the mouse pointer, users will be more inclined to notice it.

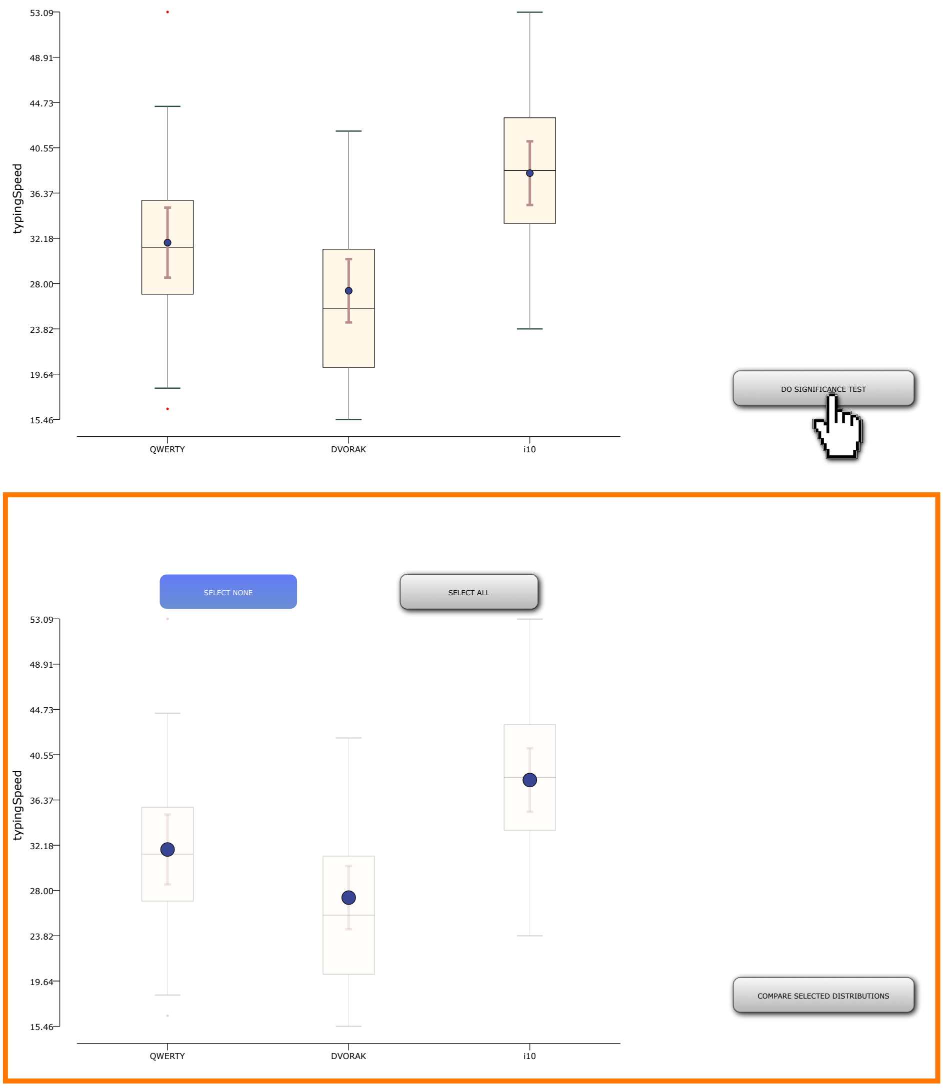


Figure 3.12: User enters a selection mode, where she can select or unselect distributions for comparison.

- ✗ This approach does not present any clear affordance. The users need to *learn* the interaction for accessing help.
- ✗ As there is no clear affordance, users are liable to hover the mouse pointers over the interface elements that do not have help.
- ✗ Some interface elements such as mean, confi-

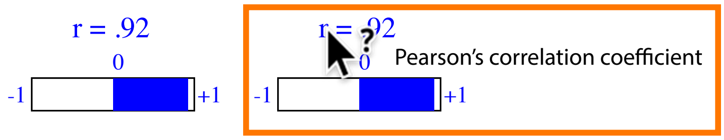


Figure 3.13: User hovers with mouse pointer over interface elements to access help.

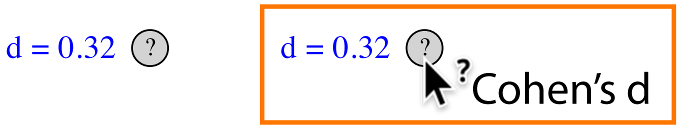


Figure 3.14: User hovers with mouse pointer over dedicated help button to access help.

dence interval, median, and so on, already have a hover-over action associated with them.

✗ This approach is not scalable (e.g., length of the text).

2. **Dedicated hover-over buttons**

   As an alternative to tooltips, we also tested dedicated hover-over buttons in our user study. Every element that has a help associated with it has a button near it. Users can hover with mouse pointer over such buttons to access help (Fig. 3.14).

Pros & Cons

✓ As the help text is shown near the mouse pointer, users will be more inclined to notice it.

✓ This approach offers a clear affordance. As a result, interface elements that have help are clearly distinguishable from those that do not.

- ✓ This approach can be used for elements that already have a hover state associated with them.
- ✗ When there are a lot of interface elements that accommodate help, the user interface is cluttered.
- ✗ It is hard to position the help button (and the help text) for all the interface elements, such as visualizations, buttons, elements of a visualization, and results of test, in a consistent manner.
- ✗ This approach is also not scalable to accommodate lengthy text.

Based on our observations from our user study, we came up with additional requirements for help:

- When users accessed help, they had a propensity to switch to an *information-gathering* mode, where they stopped performing *statistical analysis* and continued to access help for various interface elements. Therefore, there is a need for a *help mode*.
- Due to the varying length of the help text and the need for a consistent display location, a separate *display panel* is needed.

Change(s) Made

Help mode, where users hover over elements to access help

We implemented a help mode that can be toggled by a button (Fig. 3.15). To access help, user hovers over interface elements and the help text is displayed at the bottom of the page, in a display panel.

Comparison with the Old Design

- ✓ The location of help text is consistent.
- ✓ Allows users to enter an *information-gathering* mode.

Figure 3.15: Once the user is in help mode, she can hover over interactive elements to access help in the help panel.

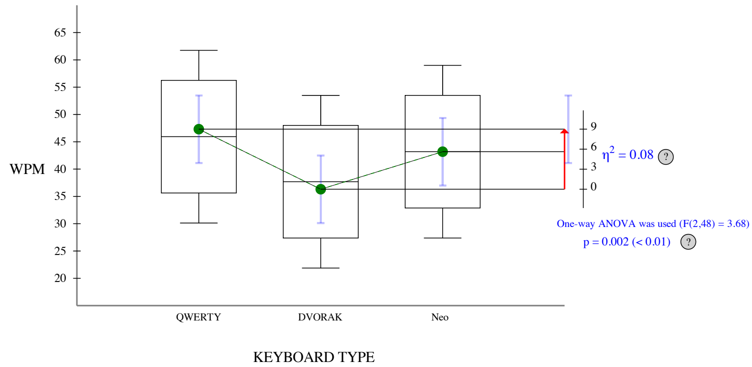


Figure 3.16: The differences between mean pairs are not shown in the old design.

✓ This approach is scalable to accommodate varying lengths of help text.

✓ It does not clutter the user interface.

✗ This approach does not offer a clear affordance in the current implementation of VisiStat. As all the elements in VisiStat have help associated with them, this is not an issue. However, for the sake of generalizability, this can be improved by fading out the interface elements that do not have help associated with them. This restricts the interaction to only those elements that have help.

3.5.3 Scale Showing the Difference in Means

Old Design

Uniform scale is used to indicate difference between means

When the user performs a significance test, the results are overlaid on the existing visualization (box plot). One of these results is the difference between the means that were compared. In the old design, a scale was used to visualize the difference between each mean pairs that were compared (Fig. 3.16).

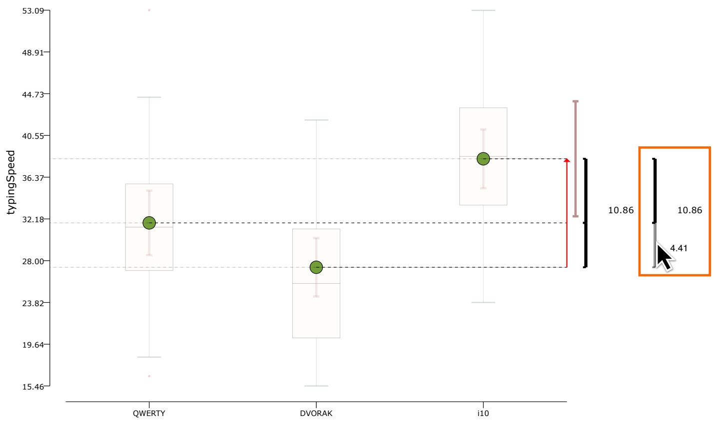


Figure 3.17: Differences between successive mean pairs are revealed progressively.

Pros & Cons

- ✗ 2 users did not interpret that the scale was in terms of typingSpeed (y-axis dependent variable).
- ✗ This design does not show the exact differences between mean pairs.

Solutions Considered	Pros & Cons
To reinforce the unit of measurement for the scale, y-axis variable can be added as a label.	–
Trace lines can be drawn from the means to intersect at the y-axis. This can be used to correctly interpret the scale.	–
The differences between every combination of mean pairs can be displayed.	✗ This approach is not scalable.

Table 3.4: Solutions that were considered for issue: scale showing the difference in means

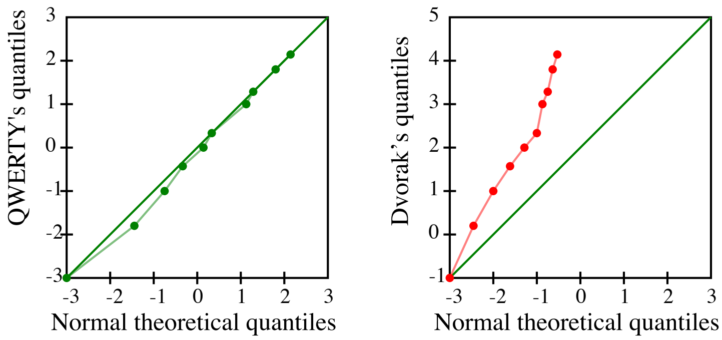


Figure 3.18: QQ-plots are used to check if a given distribution is normal or not.

Change(s) Made

Difference between pairwise means is revealed progressively

VisiStat shows only the overall difference between the compared means initially. Upon user's request, it reveals the difference between means of interest (Fig. 3.17).

Comparison with the Old Design

- ✓ This approach is scalable to several distributions.
- ✗ It reveals only the differences between adjacent means.

3.5.4 Visualization for Normality of Distributions

Old Design

QQ-plot is used to check if a distribution is normal

Normality of distributions is one of the statistical assumptions for several statistical analysis tasks. In the old design, *QQ-plots* were used to communicate this assumption to the user (Fig. 3.18). QQ-plots are used to visually compare two distributions. In a QQ-plot, to check if a distribution is normal, the quantiles of the distribution is plotted against theoretical normal distribution quantiles. If the resultant line

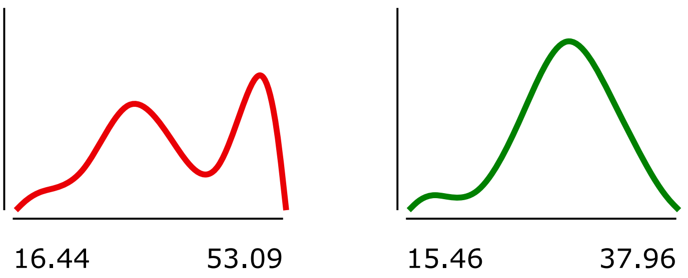


Figure 3.19: Modified histograms are used to visualize the shape of distribution, from which the user can infer the distribution type.

deviates away from a slope of 1, the distribution is probably not normal.

Pros & Cons

✓ An accurate visualization for interpreting normality of distributions (when compared to the modified histograms we present below).

✗ It is not easily interpretable for a novice user. In our user study, 6 (out of 7) users did not understand QQ-plots.

Change(s) Made

We replaced QQ-plots with modified histograms in VisiStat (Fig. 3.19). This modified visualization communicated the type of distribution by showing the shape of the distribution. Users, who are familiar with the *bell-shape* of a normal distribution, can interpret this visualization.

Modified histograms show shape of distribution

Comparison with the Old Design

- ✓ It is more easy to interpret than QQ-plots.
- ✗ It is not as accurate as QQ-plots for interpreting normality of distributions.

Chapter 4

Interaction Design

In the previous chapter, we discussed the design approach followed in VisiStat. We also discussed some issues with earlier versions of VisiStat and the solutions that were employed to fix them. In this chapter, we take a closer look at the resulting interaction design of VisiStat via a walk through. Some important features that are not part of the walk through are then described.

4.1 Preprocessing and Dataset Selection

Prior to using VisiStat, users preprocess the dataset to be in the long format (one column per variable; each row is an observation). This can be done either manually, or with the help of available tools (e.g,. DataWrangler[1], Open Refine[2]).

Users select participant column

When the dataset is loaded, users specify the column containing the ID of the participants or subjects in the experiment (Fig. 4.1). VisiStat uses this information to infer the experimental design. In our walk through, consider that our user chooses a dataset from a text entry study. Upon selection, VisiStat displays the variables in the dataset along with some background information about the experiment.

[1]http://vis.stanford.edu/wrangler/
[2]http://openrefine.org/

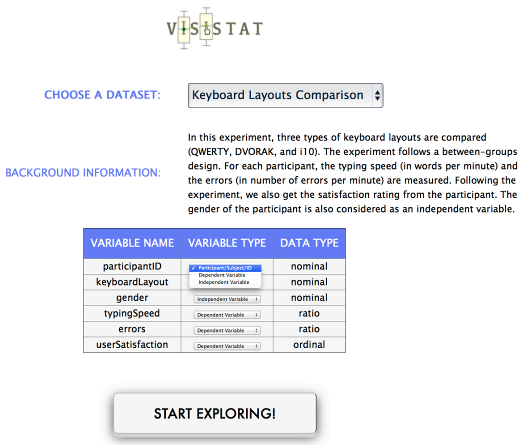


Figure 4.1: VisiStat displays some background information about the chosen dataset and allows users to specify roles for variables.

Users select variable roles

The user can also assign the role for a variable as either dependent or independent. VisiStat uses this information to select visualizations and statistical tests. Variables' roles also help the user frame research questions in EDA. Once these selections are made, the user can begin exploring the dataset.

4.2 Exploring Data

Users can change or select roles depending on research question

Once users select the dataset, they can begin exploring the data in the main page of VisiStat (Fig. 4.2). The variables in the dataset are available for selection on the left of the screen. The visualizations available in VisiStat are available at the bottom of the screen. Help button is available at the top right corner to assist novice users. The user can set a variable's role (as dependent or independent) by using a toggle button located near the variable. Although the user has the option to select the role of the variable on the pre-processing screen, users can still change the role here as the roles are subject to user's interpretation and perspective. E.g., in text entry study, typingSpeed is a dependent vari-

Figure 4.2: Variables of the dataset are displayed on the left and visualizations are displayed at the bottom

able when users consider the research question, "*Do users type significantly faster with QWERTY layout when compared to DVORAK layout?*". On the other hand, when users consider the research question, "*Does typing faster lead to more errors?*", typingSpeed is construed as an independent variable.

VisiStat uses variables' roles to select visualizations

The variables' role aids the users in framing research questions. Also, VisiStat uses this information to select appropriate visualizations. E.g., when users select an independent nominal variable, keyboardLayout, and a dependent ratio variable, typingSpeed, based on Table 3.2 "Selection of visualizations from number and type of variables", VisiStat selects a box plot of the distribution of the dependent variable corresponding to the various levels of the independent variable (Fig. 4.3).

Alternative visualizations can be used to view data in different perspectives

For a given selection of variables, alternative visualizations may be possible. These visualizations are used for reveal various properties of the data. E.g., in our walk through, the user switches to a histogram, by clicking on the button at the bottom panel, to check the data for bimodality (Fig.

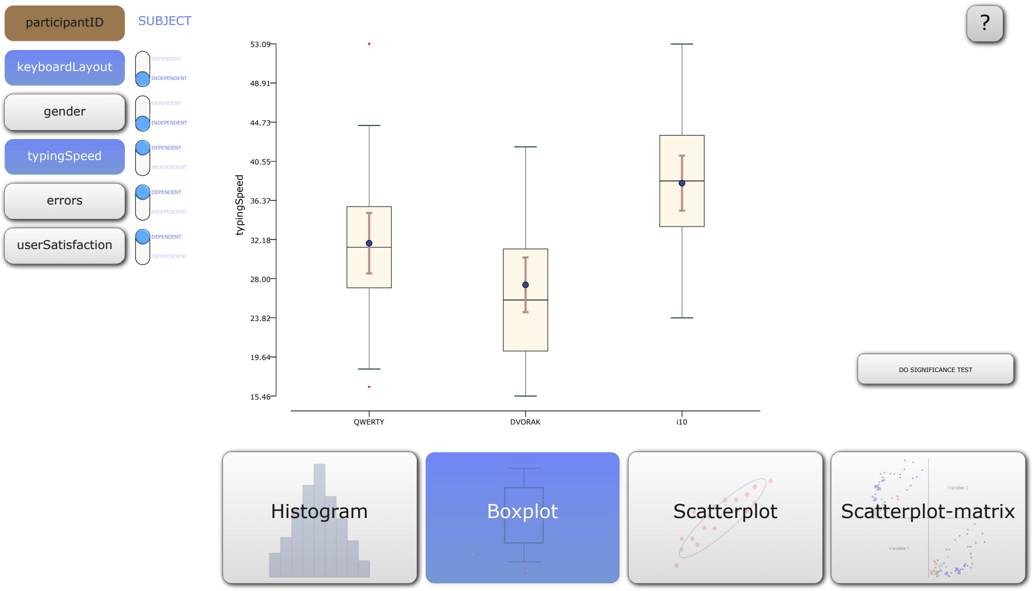


Figure 4.3: From the number and type of selected variables, VisiStat automatically chooses the visualization.

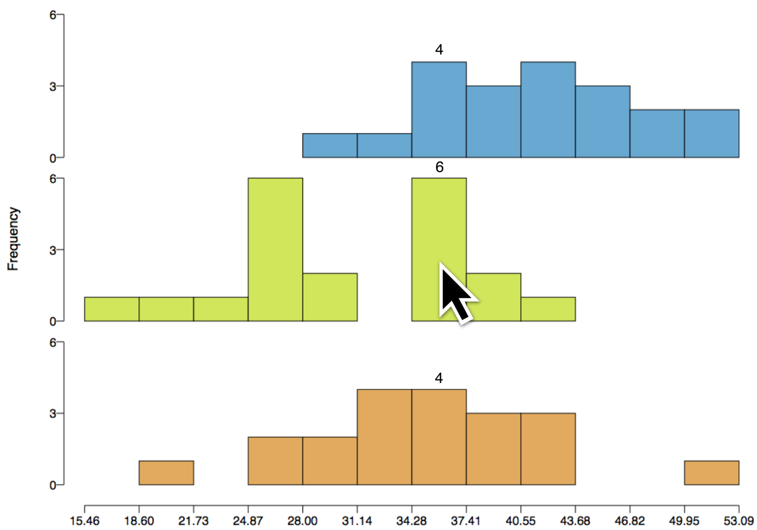


Figure 4.4: Histogram is used to check the data for bi-modality.

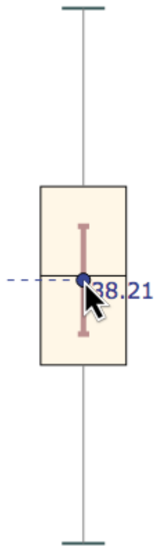


Figure 4.5: Users can hover with their mouse pointer over interactive elements to view descriptive statistics.

4.4).

4.3 Interacting with Visualizations

Users can interact with visualizations before statistical analysis

All visualizations in VisiStat are interactive and encourages users to explore data. Users can view the descriptive statistics by interacting with elements of the visualization. E.g., the user hovers with the mouse pointer over mean to view its value (Fig. 4.5). The user can use these interactions to analyze the data before performing statistical analysis. E.g., in Fig. 4.3, the user finds that the means of typingSpeed for different keyboard layouts are different. In order to test if they are significantly different, the user can perform a significance test.

To perform a significance test, the user selects the means of distributions she wants to compare (Fig. 4.6). The user can also use selection buttons for swift selection.

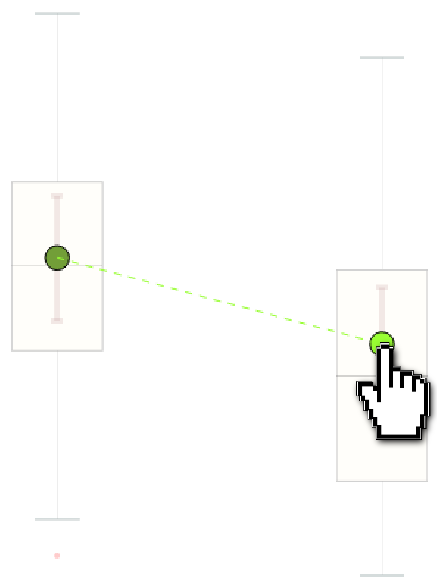


Figure 4.6: Users can selects means of distributions to compare them.

In our walk through, the user selects the means of three distributions to compare them.

4.4 Automatic Selection of Statistical Tests

VisiStat selects appropriate tests from data and prior information given by user

VisiStat determines appropriate statistical tests based on the accumulated input from the properties of data and the prior information given by the user in the preprocessing stage. In our walk through, VisiStat knows that the experiment follows a between-group design, and that the user had selected three means of keyboardLayout, a categorical independent variable. VisiStat narrows down the statistical analysis task to be performed to an One way ANOVA, a Welch's ANOVA, or a Kruskal-Wallis tests, based on the experimental design (Table A.1). The selection among these tests is based on the statistical assumptions that hold.

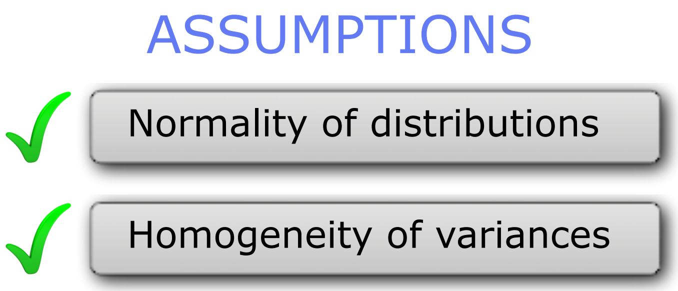


Figure 4.7: VisiStat automatically performs tests for statistical assumptions to pick the appropriate test.

4.5 Performing Assumption Tests

VisiStat knows the assumptions for statistical tasks and knows the tests to validate them

VisiStat performs tests for statistical assumptions to determine which test is to be used (parametric or non-parametric). E.g., a Shapiro-Wilk's test is used to test if a given distribution is normally distributed. Details of the test, correction procedures (e.g., transformation), and options to do alternative, non-parametric tests are made available to the user progressively. In our walk through, both the assumptions of the test, *normality of distributions* and *homogeneity of variances* are met. This information is made available to the user (Fig. 4.7).

4.6 Viewing the Results

Results are overlaid on visualization

Once the appropriate test is applied, VisiStat displays the result by overlaying it on the existing visualization (Fig. 4.8). By doing this, VisiStat builds on user's current knowledge of data, rather than altering it by separating statistical analysis from visualizations. In our walk through, since both of the assumptions are met, an One-way ANOVA is applied. We will now explore the various elements of the significance test result.

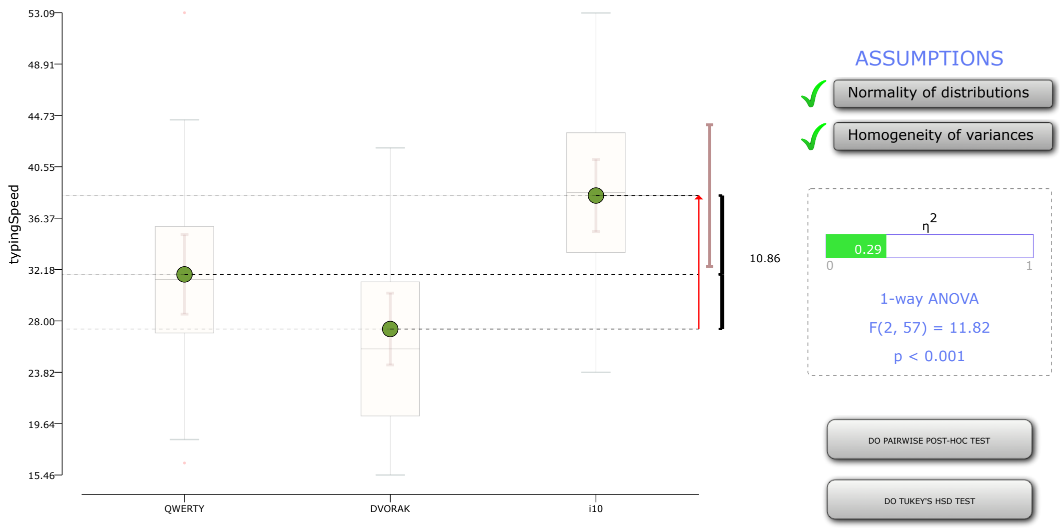


Figure 4.8: The results of One-way ANOVA are overlaid on boxplot.

4.6.1 Difference in Means

Difference in pairwise means is revealed on user's request

VisiStat displays the overall difference between the selected means. In our walk through, as one-way ANOVA is an omni-bus test, pairwise effects are not initially shown and only the overall effect of the independent variable is disclosed. The user can reveal the pairwise differences between the means by hovering with the mouse pointer over the corresponding indents on the scale in a progressive manner (Fig. 3.17).

4.6.2 Confidence Interval of the Difference in Means

The 95% confidence interval of the difference in means in also displayed (Fig. 4.9). As discussed in Section 2.1 "Problems Inherent with Statistical Analysis", VisiStat aims to gradually shift the methodology of statistical analysis from NHST to the use of effect size and confidence interval.

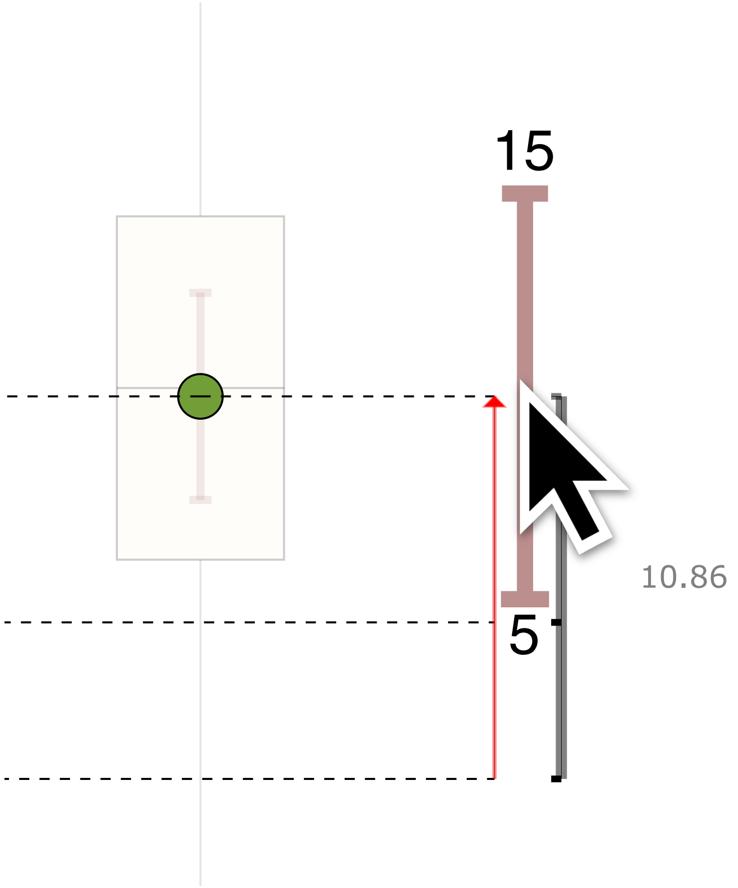


Figure 4.9: Users hover with their mouse pointer on the error bars to reveal 95% confidence intervals.

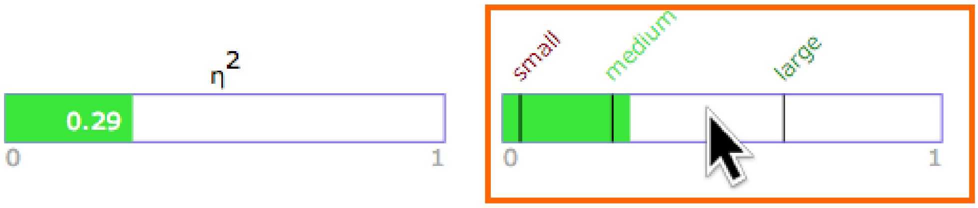


Figure 4.10: Users hover with their mouse pointer over effect size bar to interpret the magnitude.

4.6.3 Effect Size

Colors and indicators are used to interpret effect size

VisiStat visualizes the effect size at the top of the results. The value is displayed in a horizontal bar that indicates potential range of values that the effect size can take. In our walk through, the effect size is generalized eta-squared, which takes values between 0 and 1 (Fig. 4.10). In order to interpret the magnitude of the effect size, the user hovers with the mouse pointer over the bar. This information is also color-coded to depict different levels of effect size.

4.6.4 Other Statistics

Other results of the significance test such as the name of the significance test used, the test statistic, and the p-value are also displayed in that order from top to bottom (Fig. 4.8).

4.7 Further Analysis

Post-hoc tests can be done after doing initial significance tests

In most cases, the user can perform further statistical analysis tasks. E.g., in our walk through, the user can do post-hoc tests using Tukey's HSD tests or pairwise comparison using t-test or wilcox test.

If the user chooses to do pairwise post-hoc tests, she can select two means through the same interactions that were used for performing ANOVA. VisiStat checks the assumptions and applies the suitable test. The results of the test are displayed in a format similar to the One-way ANOVA, in order to maintain uniformity among significance tests (Fig. 4.11).

If the user performs Tukey's HSD test, all pairwise tests are computed and the results are visualized in a separate plot (Fig. 4.12).

The mean differences between every pairs of means are displayed with their 95% confidence intervals. The color of the

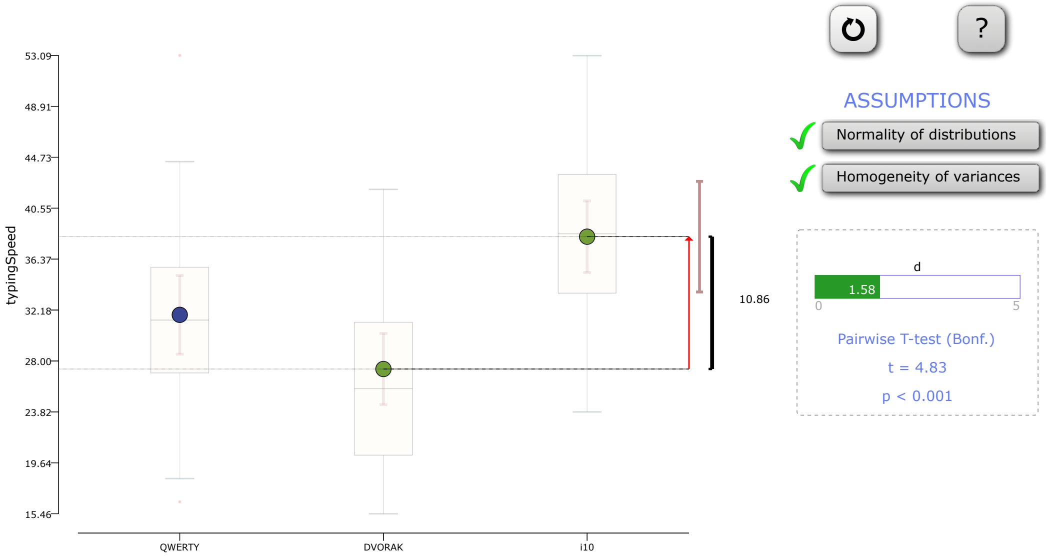


Figure 4.11: VisiStat allows users to do further analysis by performing post-hoc tests. Here, results of a pairwise t-test are shown

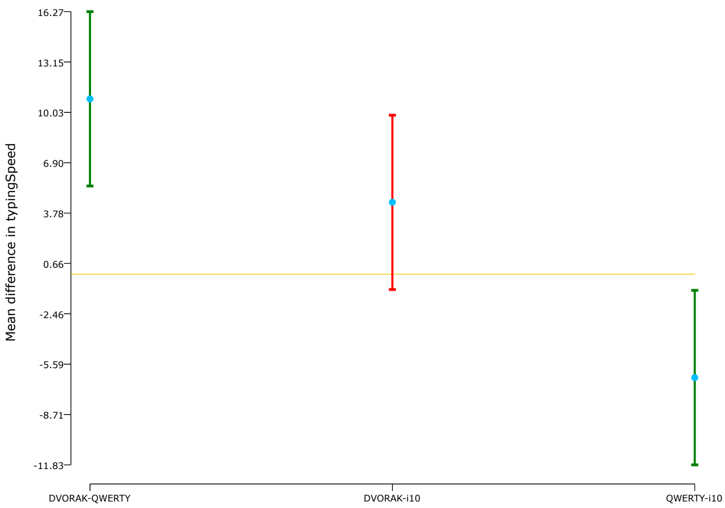


Figure 4.12: To compare all pairwise means, the user can perform a Tukey's HSD test.

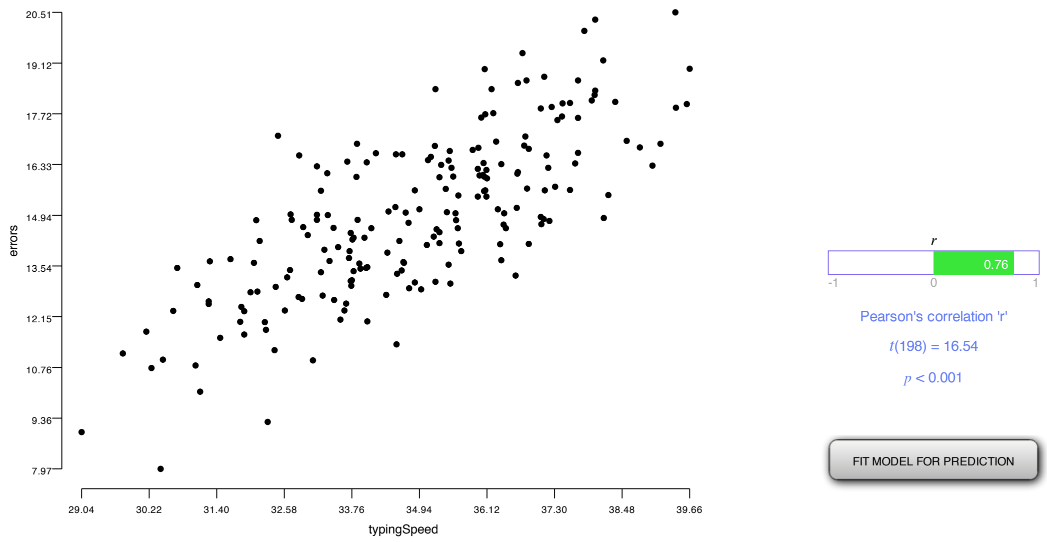


Figure 4.13: When user selects two ratio dependent variables, VisiStat overlays results of a Pearson's correlation on scatterplot.

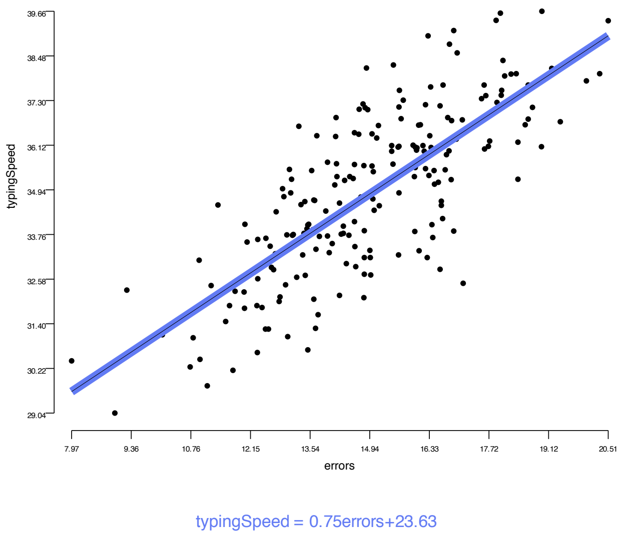
error bars indicate if they include zero or not.

4.8 Correlation and Simple Linear Regression

Users use scatterplot to find correlation

In addition to testing the effect of one or more independent variables over a dependent variable, users can analyze the relationship between two variables by using correlation. For this purpose, a scatter plot is the visualization of choice. E.g., in our walk through, the user selects two ratio dependent variables, typingSpeed and errors. A scatter plot is chosen as the default visualization and since the selected variables are ratio variables, Pearson's correlation coefficient is computed (Fig. 4.13). To further analyze the data, the user can perform regression, where the value of an outcome variable is predicted from one or more causal variables. Since regression is not commutative, VisiStat prompts the user to select the outcome variable. In our walk through, the user selects typingSpeed as the outcome variable.

Users use regression to predict values for outcome variable



errors:	10
speed:	31.13000

Figure 4.14: Users can predict the value of an outcome variable from an explanatory variable using simple linear regression

Once the user has selected the outcome variable, a linear model is constructed, which can then be used to predict the value of outcome variable from the causal variables. In our walk through, typingSpeed is predicted from errors. The linear model is then used to predict typingSpeed for an error or 10 errors per minute (Fig. 4.14).

4.9 Multiple Regression

Multiple regression is a statistical analysis task that can be used to predict the outcome variable from two or more causal variables. This is beneficial to users who want to

Users use multiple regression to predict outcome variable from multiple causal variables

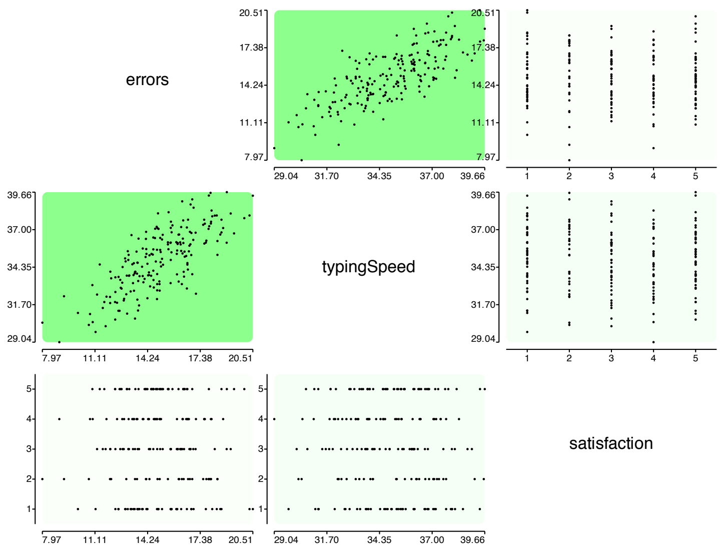
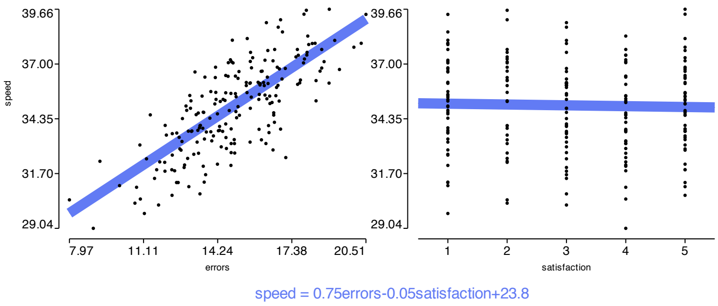


Figure 4.15: Correlation between variables is indicated by shades of green in the cells of a scatterplot matrix.

analyze the relationship between multiple variables. E.g., consider that the user wants to investigate the relationship between typingSpeed, errors, and satisfaction. Specifically, the user wants to predict the satisfaction rating of the participant from their performance, which is defined by typingSpeed and errors. When the user selects more than 2 dependent variables, VisiStat chooses scatter plot matrix as the default visualization. The correlation between each pairwise variable is indicated by different shades of green overlaid on the corresponding cell in the scatter plot matrix (Fig. 4.15). E.g., typingSpeed and errors are positively correlated and errors and userSatisfaction are negatively correlated.

To perform multiple regression, the user needs to select the outcome variable. The multiple regression model that predicts the typingSpeed from errors and userSatisfaction is visualized (Fig. 4.16).



errors:	10
satisfaction:	4
speed:	34.256

Figure 4.16: Users can predict the value of an outcome variable from more than one explanatory variables using multiple regression.

4.10 Interaction Effect

An interaction effect is said to have occurred when the effect of an independent variable on a dependent variable, differs for different levels of another independent variable. Interaction effect can have notable implications in statistical analysis. E.g., consider that the user selects the following variables: keyboardLayout, gender, and typingSpeed. typingSpeed is a dependent variable, whereas keyboardLayout and gender are independent variables. A color box plot, where the color of the plot is used to indicate the levels of the second independent variable, is chosen as the default visualization.

Users use interaction plot to check for interaction effect

When the user performs a significance test involving two independent variables, the results consist of the main effects of each of the independent variables and their interaction effect. The interaction effect is also visualized in an interaction plot, which shows the change in the dependent variable based on the levels of the independent variables.

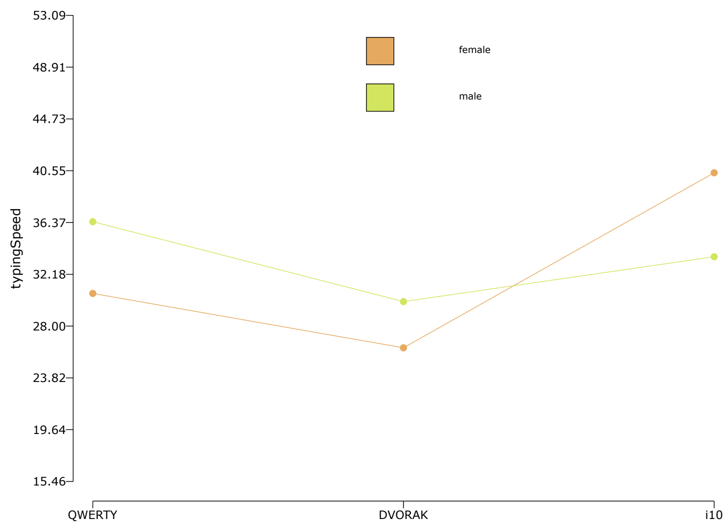


Figure 4.17: Interaction plots shows that there is an interaction between keyboardLayout and gender.

E.g., in our walk through, interaction plot shows that there is an interaction between the independent variables, keyboardLayout and gender (Fig. 4.17).

4.11 Data Transformations

Transformations are used to correct data-entry, measurement errors, and so on

Sometimes a data-entry error or a measurement error might cause statistical assumptions to be violated. To minimize such errors, VisiStat finds data transformations, which when applied to the distributions would satisfy the assumptions. The current implementation of VisiStat checks for the following transformations: log, square root, cube root, and reciprocal. When the application of one of these transformations results in the assumptions being satisfied, the transformation is automatically done. Note that, initially the user is given an option to also choose alternative tests.

Users can either transform data or do an alternative test

In our text entry study, consider that the user wants to analyze the effect of gender on the amount of errors made. To

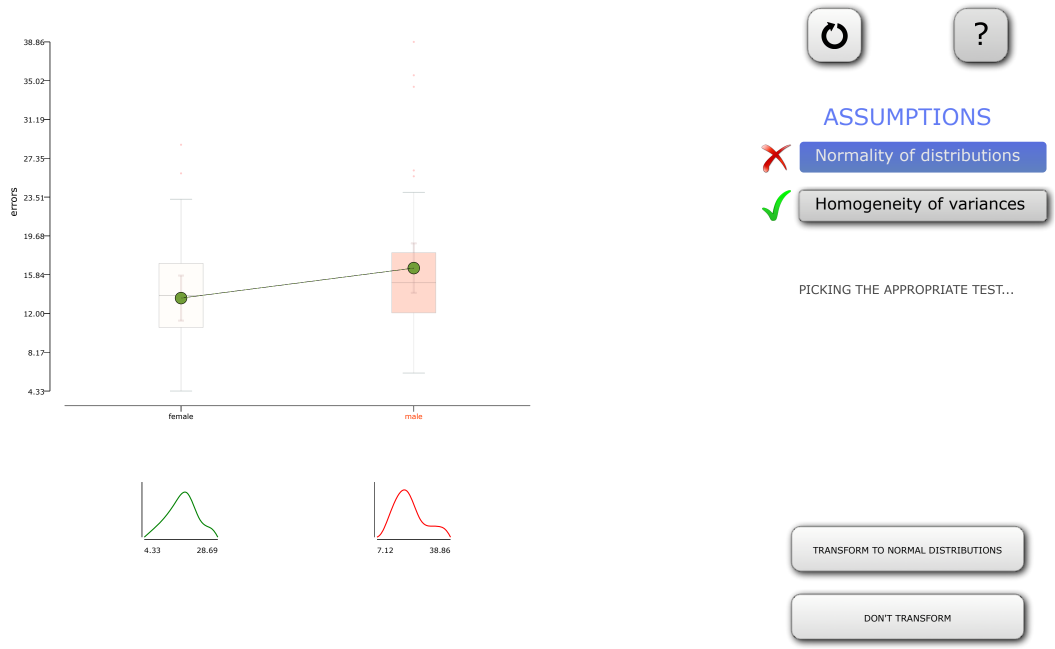


Figure 4.18: VisiStat indicates that an assumption has been violated and offers choices to do transformation or alternative test.

do this, the user selects the corresponding dependent and independent variables, error and gender. VisiStat chooses box plot as the default visualization. When the user compares the distributions of male users against female users, VisiStat performs tests for statistical assumptions and indicates that one of the distributions is not normal (Fig. 4.18).

When the user performs transformation, the distributions are transformed visually, giving direct feedback to the user. As all the assumptions for the parametric test are satisfied, VisiStat applies an unpaired t-test. Alternatively, the user can choose not to perform transformation and perform a non-parametric alternative.

Chapter 5

Evaluation

In the previous chapter, we discussed the interaction design of VisiStat. In this chapter, we discuss the user study that was conducted to assess the effectiveness of VisiStat. This chapter discusses the research questions that our user study investigates, the data collection process, and the subsequent analysis. Finally, we discuss the findings made and their implications for HCI research.

5.1 Research Questions

We investigated the following research questions in our user study:

Target users have the goal of solving certain research questions

RQ1: To which extent does VisiStat allow users to solve their research questions? Users of statistical analysis softwares typically have a clear objective of solving their research questions. E.g., a researcher who conducted an experiment that measures the typing speed of users when using different keyboard layouts would want to investigate the effect of various keyboard layouts on users' typing speed.

Many existing softwares do not take advantage of this information. As a result, users often need to think in

VisiStat allows users to solve their research questions by working with variables

terms of statistical tests and visualizations to perform statistical analysis. Since this requires an understanding of visualizations and statistical test selection procedure, such softwares have a steep learning curve and users need a large amount of prior knowledge to perform and to interpret statistical analysis (Section 2.3 "Standard Statistical Analysis Softwares"). VisiStat relieves user from this need for prior knowledge by embedding statistical knowledge in the system.

This research question investigates the effectiveness of VisiStat in helping users solve their research question. The ratio of the number of research questions solved by the user using VisiStat to the total number of research questions formulated by the user is used to quantify this.

RQ2: To which extent does VisiStat allow users to perform tasks beyond their baseline statistical knowledge?

VisiStat has potential to enable users to perform statistical analysis they have not done before

Since VisiStat embeds statistical knowledge, it can empower users to perform statistical analysis tasks beyond their baseline knowledge. This research question investigates the benefits of VisiStat to users who do not have much statistical knowledge. We do this by first obtaining a list of tasks that the user had not done before, and comparing it with the list of tasks the users managed to perform in our user study.

RQ3: To which extent does VisiStat allow users to better spot errors in statistical reports?

RQ4: To which extent does VisiStat allow users to better interpret concepts in statistical reports?

VisiStat can affect users' error spotting and interpretation skills as it is also a learning tool

In addition to the above research questions, we also wanted to investigate the effect of VisiStat on user's statistical knowledge. These research questions assesses VisiStat's potential to be used as a learning tool, in addition to allowing users to perform statistical analysis.

These skills indicate users' statistical knowledge

These research questions focus on two important skills in HCI literacy. In addition, they are also representative of the user's statistical knowledge. We measure these skills by asking users to identify errors and interpret statistical analysis concepts in a set of statistical report excerpts.

In the following section, we discuss the structure of our user study and describe the data collection process required to answer these research questions.

5.2 Structure of Final User Study

Before the user study, participants were asked to fill out a questionnaire where their level of expertise on statistical analysis tasks was gathered.

The user study was divided into two sessions that were held on successive days. On the first day, participants did a report analysis task from which their baseline error-spotting and interpretation skills are measured. The experimenter then described a dataset to them and elicited a list of research questions they are interested in solving. Participants then used VisiStat to solve these research questions. After the session on the first day, we conducted a semi-structured interview to gather qualitative feedback from the participants.

User study had two sessions to avoid learning effect

On the second day of our user study, participants performed post-test report analysis task. This measures the error-spotting and interpretation skills of the user after using VisiStat. We will now describe the various tasks involved in our user study.

5.2.1 Questionnaire

We devised a questionnaire to assess the participant's level of expertise in statistical analysis tasks used in VisiStat (Fig. B.1). For each statistical analysis task that was implemented in VisiStat (n = 20), the participant had to select his level of expertise (Table 5.1 "Table showing the options for each question in the questionnaire and its score").

Questionnaire was used to assess users' expertise on statistical analysis tasks

The results of this questionnaire were then processed and used for the following purposes:

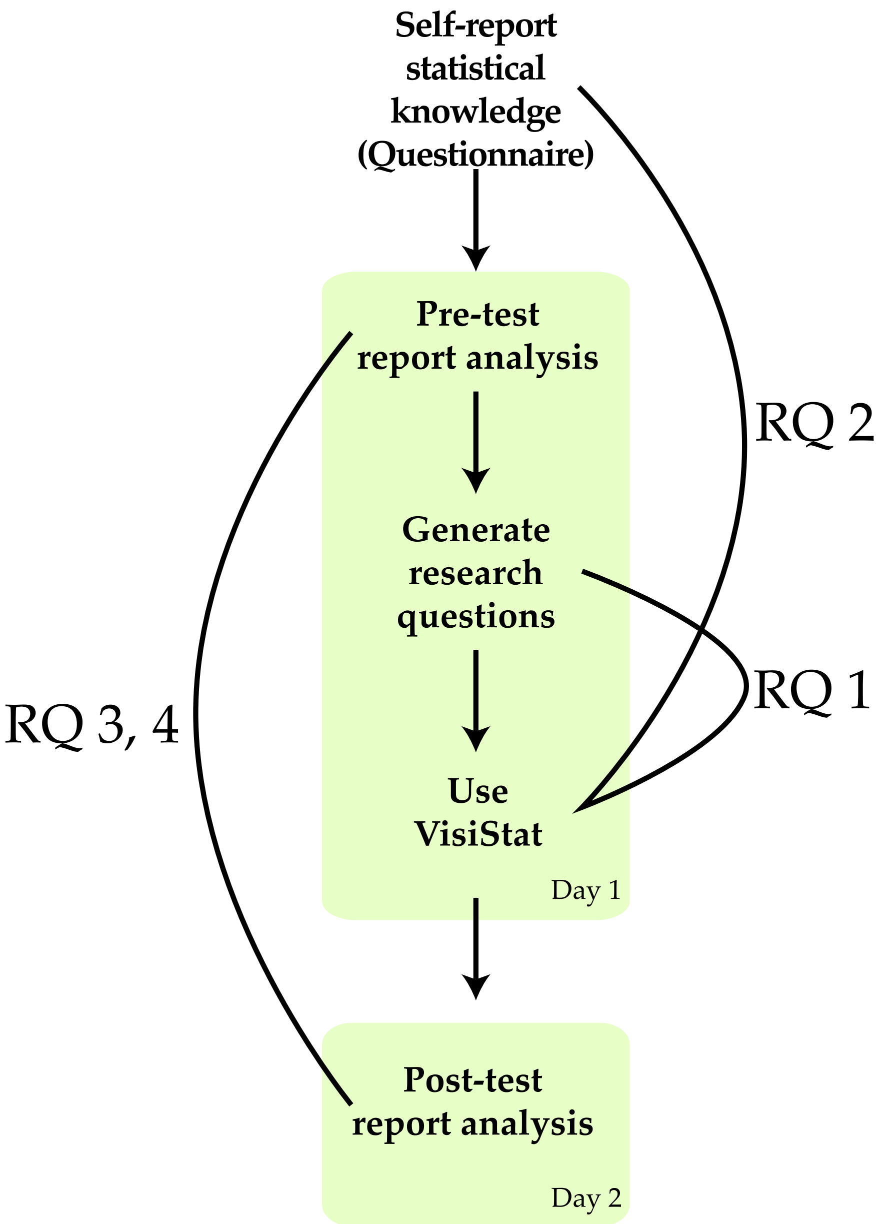


Figure 5.1: Our user study had 5 tasks spread across 2 days.

- To obtain the list of statistical analysis tasks the user had not performed before (i.e., tasks that were beyond her baseline knowledge).
- As a provisional measure of the participant's statistical knowledge. Each answer was scored and the total score for the questionnaire was used as the estimated statistical expertise of the participants (Table 5.1 "Table showing the options for each question in the questionnaire and its score"). The median of the scores were used to bisect the group of participants to novice and intermediate users (as none of the users were experts).

Option	Score
I have never heard of it	0
I have heard of it, but never used it	1
I have used it, but I am not sure about the assumptions	2
I have used it and I know the assumptions	3

Table 5.1: Table showing the options for each question in the questionnaire and its score

5.2.2 Report Analysis

Report analysis task was used to assess users' error-spotting and interpretation skills

In a report analysis task, participants were given background information about a dataset, which included a brief description of the experiment, variables involved, and their roles. Once participants were familiar with the dataset, they were asked to identify errors or missing information with excerpts of statistical reports. They were representative of statistical reports that are typically found in research papers (Table B.2 "Types of errors in the Report Analysis task with examples"). We randomized the errors in these excerpts and also made sure the number of errors were balanced between pre- and post-test report analysis tasks for all datasets (Table 5.2 "Number of errors is balanced between pre-test and post-test report analysis tasks.").

Dataset Type	Pre-test Report Analyis	Post-test Report Analysis
Between-groups design	6	7
Within-groups design	7	6

Table 5.2: Number of errors is balanced between pre-test and post-test report analysis tasks.

To minimize the learning effect, the participant could access only one excerpt at a point of time and cannot modify the answer(s) once an excerpt was handed over.

5.2.3 Qualitative Feedback

After the users used VisiStat, comments were gathered

After the participant had used VisiStat, the experimenter gathered some qualitative feedback about VisiStat. Some of the questions that were asked are:

- *"Which other statistical analysis software have you used? Can you compare the workflow between those softwares and VisiStat?"*
- *"Which feature(s) of VisiStat did you like?"*
- *"Which features did you not like or felt could have been improved? What would be your suggestions to do that?"*
- *"Would you use VisiStat in your research?"*
- *"Would you endorse VisiStat to your friends and colleagues?"*

5.3 Participants

8 participants (aged between 22 and 31, 1 female) were recruited from the local campus at the University. Participation was voluntary and they did not receive monetary compensation. The statistical expertise of the participants, assessed from the questionnaire, ranged from novice to inter-

mediate. All participants had a basic knowledge of experimental methodology and were able to generate research questions.

5.4 Results

This section discusses the data analysis process and elaborates the findings made in detail.

RQ1: To which extent does VisiStat allow users to solve their research questions?

In the task elicitation segment of our user study, the user had generated research questions of interest. The participant then used VisiStat to solve these research questions. From the audio and video logs, we gather the list of research questions that the users solved in the study. We consider only those statistical tasks that the user had interpreted for this purpose.

Users were able to answer 90% of their research questions

Participants generated a total of 51 research questions (M = 6). Of these, users used VisiStat to solve 46 (90%). Of the remainder, users had performed the appropriate statistical tests and interpreted them for 4 research questions. However, they were not confident with the results. For 3 research questions, VisiStat selected a parametric test (Mixed-design ANOVA) despite the violation of an assumption (normality of distributions), since there were no non-parametric alternatives in this case. For the other research question for which a user was not confident with the results, he construed that something was "fishy" with his dataset when an assumption was violated. As this user had the least self-assessed statistical knowledge, this finding suggests that users need basic statistical knowledge to gain VisiStat's benefits. Users were unable to solve 1 research question, as VisiStat did not have the statistical tests to do it. The research question involved finding the effect of an independent variable ($keyboardLayout$) on two dependent variables ($userSatisfaction$, $gender$). This requires the use of One-way MANOVA, which is not implemented in VisiStat.

RQ2: To which extent does VisiStat assist users to do statistical analysis tasks beyond baseline statistical knowledge?

Users performed 4 tasks they had not done before on average

The online questionnaire was used to gather the list of statistical analysis tasks that the user had not performed before (i.e., user selected options (1) or (2)). This list is indicative of the user's baseline statistical knowledge. From our user study log, we obtain the list of statistical analysis tasks each user had performed. Note that, as with the first research question, only statistical tasks that the users interpreted were considered. From these two lists, we find that users performed 32 statistical analysis tasks that they had not done before ($M = 4$). As there were different ways to solve a research question (e.g., using post-hoc test or a t-test between two distributions), we considered all such tasks that could have helped user solve a research question.

There is evidence to indicate VisiStat can vastly benefit beginners

On further inspection, we find that the amount of statistical analysis tasks done by the user beyond baseline knowledge varies from 2-8. A correlation analysis between this quantity and the self-assessed statistical expertise of the users suggests a positive correlation between them. This indicates that VisiStat is largely beneficial to beginners, who lack adequate statistical knowledge. However, this finding is not entirely convincing as the correlation is largely influenced by the extreme case of user 7. User 7, who had the least self-assessed statistical knowledge, performed the most amount of statistical analysis tasks beyond his baseline knowledge. After our user study, he commented, "*For people like me who are not too crazy about statistics, this is pretty awesome as I do not have to worry about different tests and just let the system choose the appropriate test.*" Although this indicates a dangerous potential to disregard statistical knowledge, it shows how VisiStat can benefit a beginner. However, user 3 commented that, "VisiStat reinforces the belief that what I have picked is the right test, which gives me confidence...". This shows that VisiStat boosts

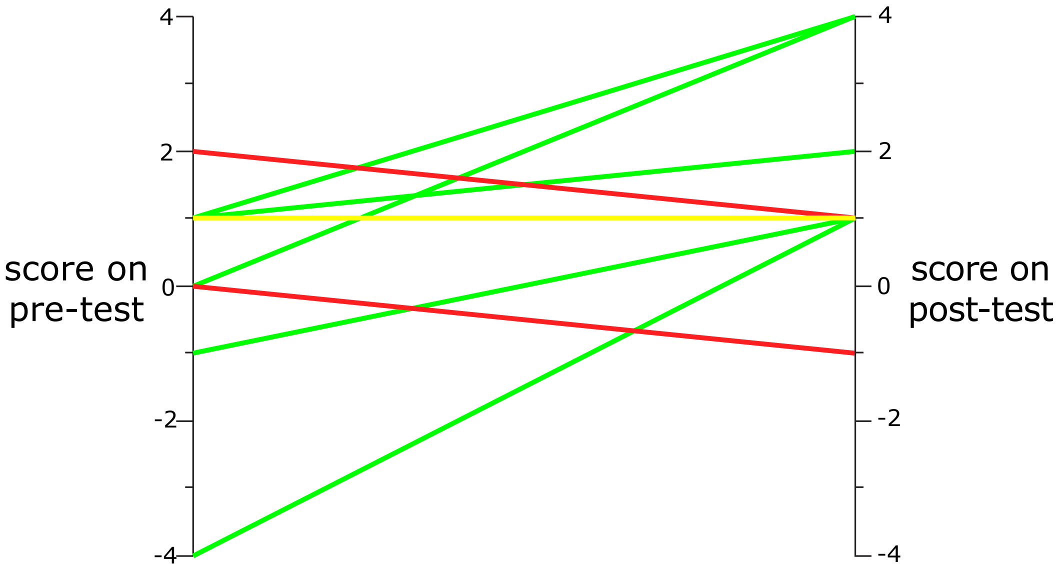


Figure 5.2: Error-spotting scores increased for 5 users, but decreased for 2 users.

confidence for users who do have some background statistical knowledge.

RQ3: To which extent does VisiStat allow users to better spot errors in statistical reports?

As we discussed earlier, error-spotting skills of the user, along with the interpretation skills reflect the statistical analysis knowledge of the user. To assess the effect of VisiStat on the error-spotting skills, we define a score,

Definition: *Score in Report Analysis*

> **SCORE IN REPORT ANALYSIS:**
> The score in report analysis is the total number of existent errors that were found by the user, subtracted by the total number of non-existent errors found.
>
> $$Score = n_{truePositives} - n_{falsePositives} \quad (5.1)$$

In the report analysis tasks, users were not given any information about scoring methodology.

Scores increased for 5 users and decreased for 2 users

We calculate the scores for pre- and post-test report analysis tasks by using the annotated excerpts. We find that the scores increased for 5 users, remained constant for 1 user, and decreased for 2 users (Fig. 5.2). However, this finding needs to be taken with a grain of salt due to the following aspects of the study:

- The sample size is too small to establish statistical significance. Mean score for report analysis increased by 1.625 (± 1.6833) from pre-test (M = 0, SD = 1.85) to post-test (M = 1.625, SD = 1.68) (paired t-test, p = 0.054, d = 1.15 (± 1.16) indicates large effect size).
- Users exhibited different levels of enthusiasm between pre- and post-test report analysis tasks. E.g., as evident from Table B.3 "Inferred behavior of participants in the final user study", users 1 and 8 were more cautious in the post-test and users 4 and 5 were more adventurous in the post-test. However, we could not find an conclusive evidence to suggest that VisiStat affected the users' behavior.

Overall, the number of true positives increased from pre-test to post-test. Also, the number of false positives decreased from pre-test to post-test.

Despite these shortcomings, the results suggest that VisiStat exhibits potential to be used as a learning tool.

RQ4: To which extent does VisiStat allow users to better interpret statistical reports?

Users comments in user study were annotated and rated by experts

To assess this research question, we annotated users' comments about various statistical concepts from the audio log of our user study. The annotations were then grouped into 5 categories, which are the elements of a statistical report.

1. p-value
2. Assumptions for the statistical test (e.g., normality of distributions, homogeneity of variances)
3. Test statistic (e.g., t statistic for t-tests, F statistic for ANOVAs)
4. Effect size (e.g., Cohen's d for t-tests, GES (generalized eta-squared) for ANOVAs)

5. Statistical test (e.g., t-test, ANOVA)

To rate these annotations, we devised *coding guidelines* (Table 5.3 "The coding guidelines for rating annotations collected from the report analysis tasks"). Each annotation can be rated as *no understanding* (rated 0), *low-level understanding* (rated 0.5), or *high-level understanding* (rated 1.0). As each statistical report involves 5 such comments, an excerpt of statistical report could have a rating in the range 0 to 5. Note that the ratings for assumptions and p-value were common for all excerpts of the report analysis task.

Coding guidelines were used to rate each comment as understood, incompletely understood, or misunderstood

We had two experts code these annotations independently (Table B.1 "Sample Annotations and Expert Ratings"). The annotations were randomized and were made anonymous to avoid bias between pre- and post-test annotations. Since the entire process of annotation, classification, and rating was time consuming, we applied our method to a sample of 1 user (46 annotations were annotated for both pre- and post-test). The rating for interpretation skills increased by 0.54 from pre-test (median = 1.25) to post-test (median = 1.5) (Wilcoxon signed-rank test was used $W = 1$, $Z = 1.98$, $p = 0.05778$, $r = 0.523$). The inter-rater agreement was found to be 0.55 (Cohen's κ). This suggests that the coding guidelines were not entirely objective. In other words, ascertaining whether a user understands a concept depended on the coder's perspective. Further analysis could be done to identify more objective coding guidelines.

Interpretation skills increased by 0.6

5.5 Limitations

One of the major limitations of our user study is the limited number of participants. It was difficult to find motivated participants, who had a sufficient knowledge of statistical analysis and experimental methodology. Additionally, participants who did volunteer for our study were not equally motivated. This is evident from the amount of time they spent in working with the system, which varied from 15m to 1h 30m.

Annotation Type	No Understanding (rated 0)	Low-level Understanding (rated 0.5)	High-level Understanding (rated 1.0)
p-value	The user either has a misunderstanding of concept or does not understand the concept.	The user knows that the p-value is used for determining whether an effect is statistically significant.	The user knows what a null hypothesis is and that p-value is used for rejecting it.
Assumptions	The user either has a misunderstanding of concept or does not understand the concept.	The user knows at least one of the assumptions.	The user knows both the assumptions.
Test Statistic	The user either has a misunderstanding of concept or does not understand the concept.	The user knows that the test statistic is used for calculating p-value.	The user knows in detail about this particular test statistic (e.g., t-statistic defines a point on the student's t-distribution).
Effect Size	The user either has a misunderstanding of concept or does not understand the concept.	The user knows that an effect size indicates the magnitude of the effect.	The user knows what the particular effect size is (e.g., Cohen's d indicates the difference between means as a factor of the pooled standard deviation).
Statistical Test	The user either has a misunderstanding of concept or does not understand the concept.	The user knows when the statistical test should be used based on experimental design.	The user knows also the assumptions of the test and alternative tests.

Table 5.3: The coding guidelines for rating annotations collected from the report analysis tasks

This affected their performance in the user study and, subsequently, the results.

Another limitation is the potential bias with the research questions that measured users' statistical knowledge (RQ 3 and 4). In-app help was used by all the participants at some point or the other, and this could have contributed to their error-spotting and interpretation skills.

5.6 Discussion

Given the lack of diligence shown by HCI researchers towards statistical analysis, the commencement of applications like VisiStat and StatWing will make statistical analysis more approachable, especially for beginners. This would result in an imminent awareness towards common statistical analysis mistakes and an increased conviction with performing statistical analysis.

Chapter 6

Summary and future work

In the previous chapter, the user study that was conducted to validate the research questions, followed by the analysis and the resultant research findings were discussed. In this chapter, the potential future works – additional features that can be added to enhance VisiStat, and a brief look at a prospective user study – are discoursed.

6.1 Summary

In the first chapter, we discussed the motivation behind VisiStat - inadequate statistical analysis expertize among HCI researchers. We tried to analyze the reasons behind this problem and saw how VisiStat mitigates the problem. VisiStat uses visualizations to help the user be aware of anomalies with data. It introduces a novel method of performing statistical analysis - by interacting with visualizations. VisiStat also reduces the amount of prior information needed by the user by embedding knowledge in the system.

In the second chapter, we discussed existing research related to VisiStat. We discussed some issues that are commonly linked with statistical analysis such as presence of outliers, shape & spread of distributions, inherent issues with Null Hypothesis Significance Testing, over-testing, inappropriate testing, disregard for statistical assumptions, and so on. We then reviewed how VisiStat solves each of these problems.

We discussed existing softwares such as TouchViz, StatWing, and ViSta and compared them against VisiStat. For use in VisiStat, we reviewed various statistical graphics and made the selection based on selection criteria.

In the third chapter, we discussed the iterative development of VisiStat by exploring various stages of its life-cycle. An initial paper prototype was assessed by expert feedback to develop a mock-up software prototype. We then discussed the user study that was conducted to find issues in interaction design with this software prototype. Some of the issues we found were: selection of means for performing a significance test, the scale for showing difference in means, in-app help, and visualization for normality of distributions. We then elaborated possible solutions these problems and highlighted the pros & cons of the changes made to VisiStat.

In this chapter, we also discussed the design principles that form the core of VisiStat: minimize prior knowledge needed by the user, allow interaction with visualizations, and reveal details on demand. We also discussed design decisions made: the approach for plotting visualizations and the scope and list of statistical tests in VisiStat.

We then took a detailed look at the interaction design in VisiStat by virtue of a walk through. We saw how VisiStat guides the users to do statistical tests and the interactions done by users to perform and interpret statistical analysis results. We also discussed other prominent features offered by VisiStat: regression, interaction plot, and data transformations.

In the chapter on Evaluation, we described the research questions that we wanted to validate. We then discussed the structure of our user study and saw how we collected data for analyzing research questions. We found that Visi-

Stat allows users to solve 90% of research questions. VisiStat also allows users to perform statistical analysis tasks beyond baseline knowledge. On average, users performed 4 tasks beyond knowledge. We also found evidence, to show that VisiStat has potential to be used as a learning tool. We found that it increases user's error-spotting and interpretation skills.

6.2 Future work

In the previous chapter, the user study that was conducted to validate the research questions, followed by the analysis and the resultant research findings were discussed. In this chapter, the potential future works – additional features that can be added to enhance VisiStat, and a brief look at a prospective user study – are discoursed.

6.2.1 Implementation

For the sake of simplicity, VisiStat had only those features that were required to work with the chosen set of datasets. For VisiStat to be used commercially available, additional features need to be added. VisiStat had a subset of statistical analysis tasks implemented. Additional tasks can easily be integrated to the system without altering the interaction design. For example, VisiStat does not have statistical tests for a numeric independent variable and this can be integrated by adding the appropriate tests. Additional features such as outlier detection and analysis, data preprocessing (possibly integrating VisiStat with an existing software [TODO: cite Data Wrangler]), clustering data points, and so on can also be consolidated to the system. Also, to improve the learn-ability of the user, dynamic learning tools such as simulations and animations can be integrated into the system.

6.2.2 User study

In order to compare the effect of VisiStat against an existing statistical analysis software, additional user studies can be conducted with two groups: participants using VisiStat as the treatment group and participants using an existing software such as StatWing or JMP as the control group. Along the same lines, VisiStat can also be assessed against other means of statistical learning such as, say, reading a book on statistical analysis, or attending a statistical analysis lecture. It could also potentially lead to interesting insights that could improve design of VisiStat.

Appendix A

Implementation

In this chapter, we discuss some of the implementation details of VisiStat.

A.1 Libraries used

VisiStat was developed in D3.js. The statistical analysis tests are computed using the statistical analysis software R. OpenCPU was used to pipeline between output of R and HTML. VisiStat is open-source and is available online on Github[1]. VisiStat is also available as an online demo[2].

Table A.1 shows the list of statistical tasks, with the ones implemented in VisiStat shown in green.

[1] https://github.com/krishna221/VisiStat
[2] http://hci.rwth-aachen.de/visistat

Number of Dependent Variables	Number of Independent Variables	Nature of Dependent Variables	Test(s)
1	0 IVs (1 population)	interval & normal	one-sample t-test
		ordinal or interval	one-sample wilcox-test
		categorical (2 categories)	binomial test
		categorical	chi-squared goodness of fit
	1 IV with 2 levels (independent groups)	interval & normal	2-sample independent t-test
		ordinal or interval	Wilcoxon Mann-Whitney test
		categorical	Chi-square test
			Fisher's exact test
	1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA
		ordinal or interval	Kruskal-Wallis test
		categorical	Chi-square test
	1 IV with 2 levels (dependent groups)	interval & normal	paired t-test
		ordinal or interval	Wilcoxon signed-rank test
		categorical	McNemar test
	1 IV with 2 or more levels (dependent groups)	interval & normal	one-way repeated measures ANOVA
		ordinal or interval	Friedman test

Continued on Next Page…

Table A.1 – Continued

Number of Dependent Variables	Number of Independent Variables	Nature of Dependent Variables	Test(s)
		categorical	Repeated measures logistic regression
	2 or more IVs (independent groups)	interval & normal	factorial ANOVA
		ordinal or interval	Ordered logistic regression
		categorical	Factorial logistic regression
	1 interval IV	interval & normal	correlation
		interval & normal	simple linear regression
		ordinal or interval	non-parametric correlation
1	1 interval IV	categorical	simple logistic regression
	1 or more interval IVs and/or 1 or more categorical IVs	interval & normal	multiple regression
			analysis of covariance
		categorical	multiple logistic regression
			discriminant analysis
2+	1 IV with 2 or more levels (independent groups)	interval & normal	one-way MANOVA
	2+	interval & normal	multivariate multiple linear regression

Continued on Next Page…

Table A.1 – Continued

Number of Dependent Variables	Number of Independent Variables	Nature of Dependent Variables	Test(s)
	0	interval & normal	factor analysis
2 sets of 2+	0	interval & normal	canonical correlation

Table A.1: List of Statistical Tasks (UCLA institute for digital research and education)

Assumption	Test
Equality of Means	Hotelling's T-square
Linearity	ANOVA test of linearity
	Correlation ratio
	Ramsey's RESET test
Unidimensionality	-
Sphericity of Distributions	Mauchly's test
Randomness	Runs test
Homogeneity of variances	Levene's test
	Brown-Forsythe test
	Bartlett's test
	White's test
Data independence	Intra-class correlation
	Graphical method
	Durbin-Watson coefficient
Independence of errors	-
Centered data	-
Normality of distributions	QQ-plot
	Kolmogrov-Smirnov test
	Shapiro-wilk test
	Back-of-the-envelope test
	D'Agostino's K-squared test
	Jarque-Bera test
	Anderson-Darling test
	Cramér-von Mises criterion
Sound measurement	Avoiding tautological correlation
	Descriptive statistics
	Avoiding attenuation
Additivity	Tukey's test for nonadditivity
Homogeneity of variance-covariance matrices	Box's M test

Table A.2: List of statistical assumptions and the tests used to validate them.

Appendix B

User Study

Annotation	Rating by Expert	Rating by Expert
"Normal distribution is when the distribution has a bell-shaped curve. Sometimes you can also use another plot to see if it is normal or not (he meant qq-plot).I haven't heard of homogeneity of variances."	0.5	0
"P-value is a probability and should be between 0 and 1. If it is less than 0.05 the result is significant."	0.5	0.5
"I don't know what a F-value is"	0	0
"I know that d is telling me how significant the effect is"	0.5	0
"T-test is used for comparing one group against another"	0.5	0.5

Table B.1: Sample Annotations and Expert Ratings

For each of the following statistical analysis task, pick the option that seems most appropriate to you. *

	I have never heard of it	I have heard of it, but haven't used it	I have used it, but I am not sure of the assumptions	I have used it and I know the assumptions
One-sample T-test				
One-sample Wilcox-test				
Unpaired two-sample T-test				
Paired two-sample T-test				
Welch's T-test				
Mann-Whitney U test				
Wilcoxon signed-rank test				
Pairwise T-test with Bonferroni correction				
Pairwise Wilcox-test with Bonferroni correction				
One-way ANOVA				
Two-way ANOVA				
One-way repeated measures ANOVA				
Mixed-design ANOVA				
Kruskal-Wallis test				
Friedman's Analysis				
Shapiro-Wilk's test				
Levene's test				
Tukey's HSD test/Tukey Krämer range test				
Pearson's correlation coefficient				
Kendall's correlation coefficient				
Biserial correlation coefficient				
Simple linear regression				
Multiple regression				

Figure B.1: Questionnaire that was used to gather users' level of expertise on various statistical analysis tasks.

Type of Error	Example
Incorrect interpretation of p-value	Male participants typed significantly faster than female participants in the typing test Unpaired t-test was used; $t(34) = 3.35$; $p = 0.9$; $d = 1.3$ (indicates large effect size)
Incorrect effect size	Male participants typed significantly faster than female participants in the typing test Unpaired t-test was used; $t(34) = 3.35$; $p = 0.02$; $\eta^2 = 0.9$ (indicates large effect size)
Incorrect statistical test	Male participants typed significantly faster than female participants in the typing test One-way ANOVA was used; $t(34) = 3.35$; $p = 0.02$; $d = 1.3$ (indicates large effect size)
Incorrect test statistic	Male participants typed significantly faster than female participants in the typing test Unpaired t-test was used; $F(34,2) = 3.35$; $p = 0.9$; $d = 1.3$ (indicates large effect size)
Incorrect p-value	Male participants typed significantly faster than female participants in the typing test Unpaired t-test was used; $t(34) = 3.35$; $p = 1.5$; $d = 1.3$ (indicates large effect size)
Incorrect interpretation of effect size	Male participants typed significantly faster than female participants in the typing test Unpaired t-test was used; $t(34) = 3.35$; $p = 0.9$; $d = 1.3$ (indicates large effect size)

Table B.2: Types of errors in the Report Analysis task with examples

User ID	Total number of errors		Inferred behavior
	Pre-test	Post-test	
1	**7**	1	More conservative
2	5	**7**	More adventurous
3	**8**	5	More conservative
4	2	**8**	More adventurous
5	1	**6**	More adventurous
6	**4**	3	More conservative
7	3	**4**	More adventurous
8	**4**	1	More conservative

Table B.3: Inferred behavior of participants in the final user study

Bibliography

John T Behrens. Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2):131, 1997.

Paul Cairns. Hci... not as it should be: inferential statistics in hci research. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*, pages 195–201. British Computer Society, 2007.

Michael J Crawley. *Statistics: an introduction using R. 327 p.* John Wiley Sons, ISBN 0-470-02297-3 (hard), 0-470-02298-1 (paper). Eine handliche und ausgezeichnete Darstellung, 2005.

Andy Field. *Discovering statistics using SPSS*. Sage publications, 2009.

Joan Garfield and Andrew Ahlgren. Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for research in Mathematics Education*, pages 44–63, 1988.

Dawn Griffiths. *Head first statistics*. " O'Reilly Media, Inc.", 2008.

Douglas H Johnson. The insignificance of statistical significance testing. *The journal of wildlife management*, pages 763–772, 1999.

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010.

Adam Perer and Ben Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 265–274. ACM, 2008.

Jeffrey M Rzeszotarski and Aniket Kittur. Touchviz:(multi) touching multivariate data. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1779–1784. ACM, 2013.

Barry Schwartz. The paradox of choice. Ecco New York, 2004.

Jenifer Tidwell. *Designing interfaces*. " O'Reilly Media, Inc.", 2010.

Edward R Tufte and PR Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.

Forrest W Young and Carla M Bann. Vista: the visual statistics system. Technical report, Technical Report 94–1 (c), 1996.

Alain F Zuur, Elena N Ieno, and Chris S Elphick. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1):3–14, 2010.

Index

abbrv, *see* abbreviation
assumptions, 65

bar chart, 19
biplot, 28
boxplot, 22

causal variable, 70, 72
Cleveland dot plot, 20
coding guidelines, 87
Cohen's Kappa, 87
confidence interval, 66
correlation, 70

data-specific knowledge, 4, 8
design principles, 34
DIA cycle, 33
dot plot, 20

effect size, 68
error-spotting skill, 78, 85
experimental research, 1
expert feedback, 33
Exploratory Data Analysis, 10

formal education, 3
future work, 93

glyph plot, 26

heat map, 29
help button, 49
help mode, 52
histogram, 19
homogeneity of variances, 65

information scent, 46
interaction effect, 27, 73
interaction plot, 27, 73

interactive visualization, 10, 35, 46, 54, 63, 66, 68
interpretation skill, 78, 86
issues with statistics, 8
iterative design, 33

Kappa, 87

learning effect, 79
letter-value summary, 22
limitations, 87
linear regression, 70

Many Eyes, 11
modality, 61
multiple regression, 71

normality of distributions, 65
Null Hypothesis Significance Testing, 8

outcome variable, 70, 72

paper prototype, 33
participants, 45, 82
pie chart, 29
plotting approach, 36
post-hoc test, 68
preprocessing tools, 59
prior knowledge, 3, 12, 14, 35, 65
progressive disclosure, 36, 56

QQ-plot, 56
questionnaire, 79

rating criteria, 87
report analysis, 81
research hypothesis, 77, 83
research questions, 77, 83

scatterplot, 24, 70
scatterplot matrix, 24, 71
score, 85
sparkline, 30
star plot, 26
Statistical analysis softwares, 12
statistical knowledge, 3, 12, 14, 35, 65, 79, 84
StatWing, 14
stem-and-leaf display, 21

Tableau, 11
ternary chart, 31
think-aloud protocol, 33

tooltips, 49
TouchViz, 10
transformations, 74

user study, structure, 40, 79

variable role, 60
ViSta, 15

wilcoxon signed-rank test, 87
Wilkinson dot plot, 20

Typeset March 10, 2014