

# VisiStat: Visualization-driven, Interactive Statistical Analysis

Krishna Subramanian  
RWTH Aachen University  
52062 Aachen, Germany  
krishna.prasath.subramanian@rwth-aachen.de

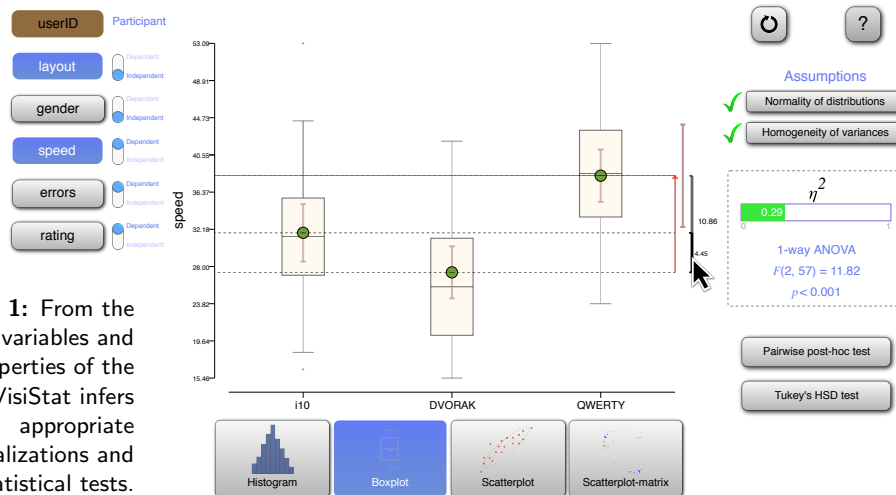


Figure 1: From the selected variables and properties of the data, VisiStat infers appropriate visualizations and statistical tests.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).  
CHI'14, April 26–May 1, 2014, Toronto, Canada.  
ACM 978-1-4503-2474-8/14/04.  
<http://dx.doi.org/10.1145/2559206.2579423>

## Abstract

To promote the practice of sound statistical analysis in HCI, we introduce VisiStat, a tool that allows users to perform statistical analysis by interacting with visualizations. It guides users to select the appropriate statistical analysis tasks based on the research questions they want to answer. By collocating statistical analysis results with appropriate visualizations, users are made aware of data-specific knowledge, which consequently improves their understanding of data and reduces common statistical analysis mistakes. In our user study, VisiStat helped users to answer 90% of the research questions they posed. On average, the users performed four statistical analysis tasks beyond their prior experience.

## Author Keywords

Statistical analysis; progressive disclosure; interactive data visualization

## ACM Classification Keywords

G3 [Probability and Statistics]: Statistical software; H.5.2. [Graphical user interfaces (GUI)]: .

## Introduction

Statistical analysis is an integral part of HCI research. However, there is cause for concern over the lack of adequate statistical analysis knowledge among HCI

**Table 1:** Knowledge involved in statistical analysis

**Statistical knowledge:**

- This includes information such as selection procedure of statistical tests, list of assumptions for these tests, and interpretation of various terms in the result.
- This knowledge is a major barrier for HCI researchers who lack a strong mathematical background.
- Books, online courses, and seminars attempt to bridge this gap [3, 4, 6]. However, these require major learning efforts.

**Data-specific knowledge:**

- This includes information such as shape of the distribution, awareness to outliers, and transformation of data.
- Data-specific knowledge is updated during the course of the analysis, and data visualizations often help in revealing and reminding users of this knowledge.

researchers. A survey conducted on research papers in leading HCI journals revealed that 40 of 41 sampled papers had issues with the reported statistics [1]. Some of the prominent issues were inappropriate testing, over-testing, and a disregard for checking assumptions. The immediate effect of this is that it invalidates, or at the least obfuscates, the author's findings.

Statistical analysis requires both statistical knowledge and data-specific knowledge to be *in the user's head* (Table 1). Current statistical analysis software often separates data visualization from statistical analysis results (e.g., *p* values or effect sizes). The latter are often presented in an isolated table. This separation between visualization and statistical results could lead to the analyst overlooking potential anomalies in the data (e.g., outliers), resulting in misinterpretations and, subsequently, mistakes with statistical analysis.

VisiStat attempts to improve the quality of statistical analysis in HCI research in two ways: (1) it guides the user to select the right statistical analysis procedure by embedding the statistical knowledge in the system, and (2) it collocates the statistical results with appropriate visualizations to remind users of data-specific knowledge.

**Related work**

Previous research in the field of statistical analysis has dealt with interactive visualization of raw data and automated statistical procedure selection.

TouchViz is a visualization tool for tablets that has a rich, interactive design [5]. It uses multi-touch gestures based on physical metaphors to enable the user to perform rudimentary statistical analysis tasks such as clustering and filtering. However, it is limited to descriptive analyses, such as clustering and filtering.

StatWing<sup>1</sup> is a web-based, commercial software that allows users to perform statistical tests. Like VisiStat, StatWing automatically selects the appropriate statistical procedure based on research questions and augments the result with visualizations. However, StatWing displays graphs *as a result* of the analysis. Users may overlook dataset-specific knowledge, apparent in the visualization, because of premature conclusion. The lack of data visualizations prior to statistical analysis can cause users to overlook dataset-specific knowledge that may be apparent in the visualization. To prevent this potential negligence, in VisiStat, users *initiate* statistical analysis by interacting with visualizations.

In the following sections, we present the design principles employed in VisiStat followed by a walkthrough of the interaction and the user study.

**Design Principles**

VisiStat enables users to perform complex statistical analyses by placing both statistical knowledge and data-specific knowledge upfront. This goal is elaborated in the following design principles:

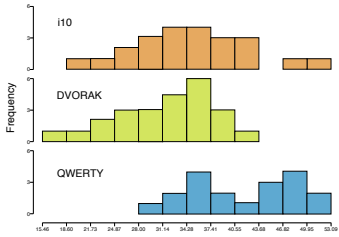
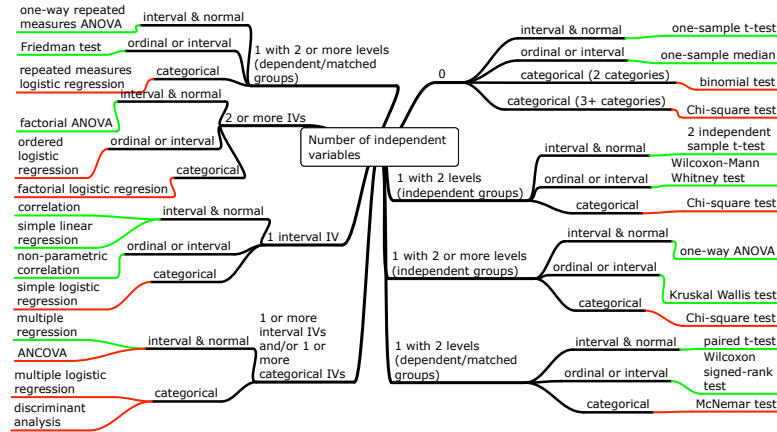
**Embedded statistical procedure decisions:** During the preprocessing stage, users specify the type and role of each variable in the dataset. This information is used by VisiStat to select the appropriate visualization and statistical procedures. This relieves the user from the need for prior extensive statistical knowledge.

**Interactive visualizations:** Users initiate statistical analysis by interacting with visualizations. This enables them to identify potential anomalies in the data before statistical analysis. Also, as seen in TouchViz, interactivity

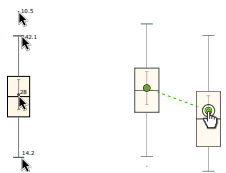
<sup>1</sup><http://statwing.com>

encourages users to perform Exploratory Data Analysis [5].

**Figure 2:** A decision tree showing tests in VisiStat by the number of independent variables and by the type of dependent variables



**Figure 3:** Histogram mode is used to check data for bimodality.



**Figure 4:** Hovering over a mean displays its numerical value. The user selects means for comparison by clicking on them.

**Progressive disclosure:** The details of statistical assumption tests and descriptive statistics are shown on demand. E.g., an indication of whether each assumption has been satisfied or violated is revealed initially. When a violation occurs, or upon user request, detailed results of the assumption accompanied by additional visualizations are shown.

### Interaction Walkthrough

Here, we describe a typical workflow of VisiStat. Prior to analysis, users preprocess data to be in long format<sup>2</sup> manually or with the help of available preprocessing tools such as [2]. Users also specify the column containing User IDs of participants in the experiment, from which VisiStat infers the experimental design. For this walkthrough, we use a dataset from a text entry study that follows a between-subject design.

<sup>2</sup>One column per variable; each row is an observation

### Inferring visualization based on variable selections

The main screen is shown in Fig. 1. Users can select roles of the variables (dependent or independent) on the left. This aids the user in framing research questions. Also, based on the variable selection, VisiStat selects the visualizations that are most likely to lead to further analyses. Here, *keyboardLayout* (categorical independent variable) and *typingSpeed* (interval dependent variable) are selected. VisiStat shows a boxplot of the typing speed corresponding to each layout.

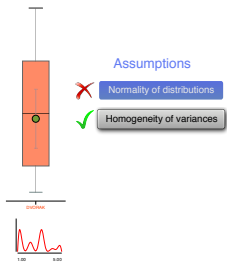
At the bottom of the screen, users can choose other possible visualizations for the selected variables to reveal different aspects of the data. From the previous example, the user may want to check if the data is multimodal by switching to histogram mode (Fig. 3).

### Interacting with visualizations

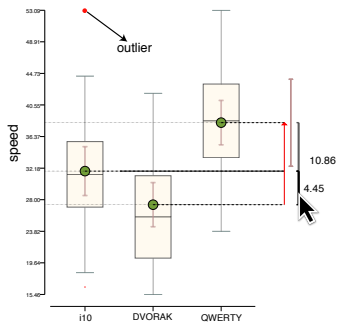
Elements of the shown graphs, such as bar, box, or whiskers, are interactive, allowing descriptive statistics to be revealed on-demand, by hovering with the mouse pointer over bands and whiskers of boxplots (Fig. 4). Users also initiate statistical tests through these elements. In this walkthrough, the user has selected the means of the distributions to compare them.

### Automatically selecting statistical tests

VisiStat determines appropriate statistical tests based on current visualization, type of selected variables, and their roles. Here, the dataset follows a between-subject design, and the user had selected three means of *layout*, which is a categorical independent variable. VisiStat infers that the user wants to compare means, and used One-way ANOVA, which is appropriate for the selected variables. Currently, VisiStat implements common tests for interval-dependent variables (Fig. 2).



**Figure 5:** Violation of the normality assumption leads to further diagnostic plots.



**Figure 6:** The difference between means are visualized next to the data. Hovering over the brackets shows pairwise differences.



**Figure 7:** Effect-size magnitude and range is visualized in a bar. Hovering over the bar shows legends assisting interpretations.

### Performing necessary assumption tests

VisiStat automatically performs necessary tests of statistical assumptions. Here, normality of distributions and homogeneity of variances are tested. Details of the tests, correction procedures (e.g., transformation), or alternative nonparametric tests are disclosed to the user progressively (Fig. 5).

### Viewing results

Besides the numerical results of statistical analyses, an interactive layer pertaining to the results augments the visualization. In Figure 6, the difference between consecutive pairwise means overlays the ANOVA results. Hovering over each pair reveals the difference between the corresponding means.

VisiStat visualizes some of the numerical results to help in interpretation. In Figure 7, a color-coded bar is used to visualize the magnitude and the possible range of  $\eta^2$ .

VisiStat overlays the results on the graphs to visually remind the user of data-specific knowledge. In Figure 6, outliers can be spotted in the boxplot and be highlighted after mean comparison tests.

### Performing further analyses

Users can perform further analyses by interacting with visualizations. E.g., the user can select two means in the mean comparison results for a pairwise post-hoc test. VisiStat determines an appropriate test based on the selected data. E.g., selecting two normally-distributed means invokes a pairwise  $t$ -test, with non-parametric alternative of a pairwise wilcox-test being invoked when distributions are not normal.

## User Study

We conducted a user study to investigate whether VisiStat aids users to perform statistical analysis tasks and also to explore VisiStat's influence on users' statistics skills. Our research questions were:

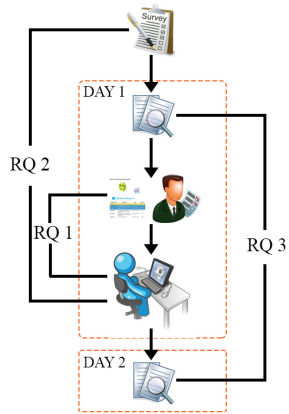
- RQ1 To which extent does VisiStat help the users to perform the statistical analysis tasks that they want?
- RQ2 To which extent does VisiStat guide the users to perform statistical analysis tasks beyond their prior statistical knowledge?
- RQ3 To which extent does VisiStat enable users to better spot errors in statistical reports?

These research questions deal with statistical knowledge that cannot be directly measured. In this study, we used observable indicators to represent statistical knowledge as will be discussed in the Results section.

### Procedure

The structure of our user study is shown in Figure 8. Prior to the user study, our users filled out an *online questionnaire*, where they self-assessed their expertise on various statistical analysis tasks. The lab study consists of two sessions held on two consecutive days. On the first day, users performed a *pre-test* error-spotting task and worked with VisiStat. On the second day, they performed a *post-test* error-spotting task to assess the difference in statistics skills.

In each *error-spotting task*, users received metadata about one dataset. This included a short description and the motivation of the experiment, the experimental design, and the list of controlled/measured variables and their



**Figure 8:** Steps in our user study and the data used to answer each of our research questions.

Users typed significantly faster with QWERTY layout when compared to DVORAK layout.

Unpaired t-test was used

$t(34) = 3.35$

$p = 0.001$

$d = 1.3$  indicates large effect size

$p$ -value is too high for significance

**Figure 9:** An excerpt shown in the error-spotting task. The error is annotated in red.

data types. User were then asked to read 7 excerpts of statistical reports pertaining to the given dataset (Fig 9). They had to identify statistical errors and ambiguous information that could jeopardize each excerpt. For each identified error, the experimenter ascertained the users' reasoning by asking followup questions. The number of errors were balanced between the set of excerpts used for pre-test and post-test error-spotting tasks.

For each VisiStat task, users were introduced to a different dataset and were asked to note down research questions that they were interested in exploring. Then, they used VisiStat to explore the data and answer the research questions in a think-aloud session.

We recorded the usage log, screen, and audio. After the session, we conducted a semi-structured interview for further comments about the user experience. Average duration on the first day was 1h 15mins for each user.

In the post-test *error-spotting task*, users were tested with the same dataset as in the pre-test, but with a different set of excerpts. The post-test was done on the second day to avoid short-term retention.

## Participants

Eight volunteers (aged between 22-31; 1 female), who were students and researchers with an HCI background, were recruited from the local university to participate in our study. All participants had knowledge of experimental study design. They rated their statistical expertise from novice to intermediate.

## Results

### Users can do their planned analyses with VisiStat

From the recordings we identified whether users analyzed the research questions they had planned, and how

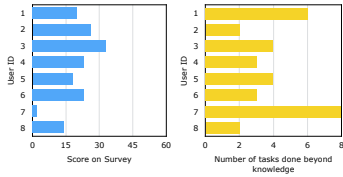
confident they were with their analyses. Users identified 51 research questions ( $M = 6$  questions/user). Of these, users correctly analyzed and interpreted 46 questions using VisiStat. There were four questions for which our users had correct results but were not confident, as VisiStat had applied a parametric test despite the violation of an assumption (because there were no alternative tests in this situation). This suggests that the presence of assumption indicators made the users be aware of this violation and, thereby, be more cautious with their interpretation. On average, users came up with 6 research questions and were able to answer 90% of them confidently using VisiStat.

### VisiStat enables users to perform statistical tests beyond their previous knowledge

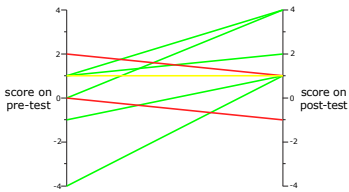
For the formulated research questions, users needed to do on average 7.5 statistical tests beyond their self-reported knowledge. Of these, VisiStat guided them to perform on average 4.25 statistical tests. In an extreme case, User 7, who self-rated statistical knowledge very low, performed most tests beyond previous knowledge (Fig. 10). However, his comment indicates a dangerous potential to be overdependent on VisiStat: "... I don't have to worry about the different tests and just let the system choose the appropriate test". Nevertheless, User 3 remarked that VisiStat improves his conviction on performing analysis: "VisiStat reinforces the belief that what I have picked is the right test, which gives me confidence ...".

### Influence on error-spotting skills is inconclusive

Users' utterances while identifying errors were annotated and classified into various types based on the statistical terms (e.g.,  $p$ -value, test statistics, and the significance test used) and the type of error (e.g., incorrect value and incorrect interpretation). We collected both the true-positives (correctly identified errors) and the



**Figure 10:** Users' scores in the online questionnaire and the number of tasks completed beyond prior knowledge.



**Figure 11:** Change in scores for all users between pre- and post-test error-spotting.

false-positives (identified errors that did not exist). The differences in the quantity and quality of error description indicates the knowledge that participants acquired from interacting with VisiStat. Overall, the number of true-positives increased from 17 to 24, while false positives decreased from 17 to 11. The score in the error-spotting task is calculated by subtracting the number of false positives from the number of true positives. On comparison between the pre- and post-tests, the score increased for five users, remained constant for one user, and decreased for two users (Fig. 11).

### Limitations

We had a limited number of users for the study and they were not equally motivated. This influenced the amount of time they spent working with the systems (from 15 min to 1.5 hr) and, subsequently, their performance in the user study. In the error-spotting task, behavior changed from pre-test to post-test. Some users were more adventurous and tried to find more errors, whereas some users were more cautious and exercised restraint. As a result, the lack of uniformity in user involvement affected their score. These confounding factors contribute to the lack of consistent patterns of how VisiStat influenced their behavior.

### Future Work

We plan to investigate how VisiStat complements other modes of learning, e.g., video lectures, self-learning materials [6], and interactive statistics tutorials<sup>3</sup>. We also plan to include more statistical analysis procedures in VisiStat, e.g., statistical tests for categorical dependent variables.

<sup>3</sup>E.g., <http://swirlstats.com/>

### Conclusion

We presented VisiStat, an interactive data analysis approach that infers visualization and statistical tests based on users' interactive input. Our user study shows that VisiStat (1) allows users to find answers to their research questions, and (2) assists them to perform statistical analysis tasks beyond their baseline statistical knowledge. We envision that VisiStat will improve the quality of statistical analysis in HCI community.

VisiStat is open-source. An online demo and source code are available at <http://hci.rwth-aachen.de/visistat>.

### Acknowledgements

This work was funded in part by the German B-IT Foundation.

### References

- [1] P. Cairns. HCI... Not As It Should Be: Inferential Statistics in HCI Research. In *BCS-HCI '07*, 195–201.
- [2] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer. Wrangler: Interactive Visual Specification of Data Transformation Scripts. In *CHI '11*, 3363–3372.
- [3] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. Wiley Publishing, 2010.
- [4] H. C. Purchase. *Experimental human-computer interaction: a practical guide with visual examples*. Cambridge University Press, 2012.
- [5] J. M. Rzeszotarski and A. Kittur. TouchViz: (Multi)Touching Multivariate Data. In *CHI EA '13*, 1779–1784.
- [6] J. O. Wobbrock. Practical statistics for human-computer interaction. In *Annual Workshop of the HCI Consortium (HCIC '11)*.