Extending Spreadsheets to Support Seamless Navigation at Scale

Sajjadur Rahman, Mangesh Bendre, Pingjing Yang, Yuyang Liu, Shichu Zhu, Zhaoyuan Su,
Kevin Chen-Chuan Chang, Karrie Karahalios, Aditya Parameswaran

Department of Computer Science, University of Illinois Urbana-Champaign

Abstract

Spreadsheets are one of the most popular tools for ad-hoc exploration and analysis of data. As spreadsheets stretch to support larger and larger datasets, operating on these large datasets via a spreadsheet interface can be overwhelming for end-users. In particular, the absence of an overview and a lack of context can cause difficulty in navigation and manipulation as users interact with large collections of data in spreadsheets. In this paper, we propose the integration of an overview interface with spreadsheets to support a seamless exploration of large datasets by enabling modes of operation commonly found in online maps, e.g., panning, semantic zooming, brushing and linking. We develop NOAH*, an in-situ zoomable overview interface for spreadsheets that allows users to navigate and analyze spreadsheet data at various granularities.* NOAH *preserves the spreadsheet semantics, and look and feel, while also allowing users to navigate seamlessly, and gain a high-level or low-level perspective of the spreadsheet data. We conducted a user study which demonstrated that* NOAH *made it easier for the participants to navigate the spreadsheet data relative to Microsoft Excel as they performed different tasks. The study showed that spreadsheet tasks involving navigation that were hard to do correctly or complete in a timely manner in Excel, were easy to complete correctly and quickly (≈ 2X faster) in* NOAH.

1. Introduction

With a user base of more than one-tenth of the world's population, spreadsheets are one of the most popular tools for ad-hoc exploration and analysis of data [exc]. Spreadsheets enable users to view, structure, and present data in an intuitive tabular layout, wherein users can map their data and tasks. This tabular layout is one of the primary reasons for the popularity of spreadsheets [NM90].

As spreadsheets continue to support increasingly large scale data, *e.g.*, Google Sheets currently supports two million cells, a five-fold increase from the previous limit of 400K cells, "navigating" the data within this the tabular layout becomes extremely difficult. Navigation is the process of viewing and manipulating the computer display to show another portion of the information space [BI90]. In this paper, we focus on three fundamental spreadsheet operations [BAM14, LBPFJ09] that are impacted due to difficulties in navigation, namely *(a)* scrolling, *(b)* steering, and *(c)* analyzing data. Even with relatively smaller data, navigation difficulties impacted these operations as discussed below.

Scrolling Challenges. Spreadsheets adopt a scrolling approach to pan the viewport to display different portions of a spreadsheet data at a time. However, this approach introduces a discontinuity between the displayed information within the spreadsheet. In particular, without a bird's eye view of the overall structure of the spreadsheet, users need to mentally assimilate the layout of the data as they navigate via scrolling and in the process can easily lose context of where they are and where they should go next [WPW99]. This loss of context can cause cognitive burdens for the users [CKB09a], especially on larger datasets. Even with small spreadsheets, users add landmarks, *e.g.*, headers colored cells, as visual cues to assist in navigation [WPW99]. Some users take even more drastic measures to avoid getting lost during navigation, *e.g.*, sketching maps of spreadsheets on paper [WPW99].

Steering Challenges. Steering tasks in spreadsheets involve users dragging the mouse pointer through a spreadsheet spanning multiple viewports. For example, to compute a conditional count formula, `COUNTIF`, on a subset of data with thousands of rows, a user needs to first tediously steer (drag) the mouse pointer through the appropriate spreadsheet cells to select the data. However, steering tasks are often time consuming and can lead to errors in formula, *e.g.*, selecting an incorrect subset of cells.

Analysis Challenges. Analysis tasks in spreadsheets require planning and puts cognitive burden on the users [KMB03]. The problem is further exacerbated by the visual discontinuity introduced due to the limited viewport visible to the user. For example, the user may be interested in comparing spatially separated subsets of data within the spreadsheet, forcing the user to move back and forth between multiple viewports, overwhelming the user in the process. As an alternative, users tend to copy subsets of data side by side to reduce the visual discontinuity [NM90, WPW99], which is cumbersome.

Prior Work. Commercial spreadsheet software, *e.g.*, Microsoft Excel, address the aforementioned navigational challenges to varying degrees of success. For example, to maintain the context of navigation, Excel allows users to manually create landmarks via named ranges [nam], which are essentially references to a region within the spreadsheet, and appear in a *name box* located to the left of the formula bar. Users can select a named range to navigate to the referred region. However, the onus is on the user to create the named ranges for every single potential interaction. Excel provides keyboard shortcuts to select entire ranges of data, as a primitive solution to steering. However, for selecting a subset of cells, users need to resort to dragging via the mouse pointer. Excel provides pivot tables to summarize the spreadsheet data via aggregation. The summarized view enables users to compare subsets of data without having to move back and forth between the subsets, thus reducing visual discontinuity. However, as pivot tables do not exist in-situ with data, users are unable to view and access the data underlying the summary for clarity as well as verification, a desirable feature of any summarized view [M*17]. In summary, there is a lack of overview interfaces for spreadsheets that can address the user's navigation challenges related to scrolling, steering, and analysis of data in-situ.

Open Questions. Building such interface, however, offers a host of challenges spanning from interface usability to system architecture. What are the design considerations for developing an overview interface that addresses the navigation challenges? How do we seamlessly integrate the overview interface with spreadsheets without disrupting the spreadsheet functionalities? What is a suitable data structure to support such an integration so that the overview and spreadsheet remain interoperable? How do we ensure that the interface is applicable to all the existing spreadsheets that support a variety of data types?

NOAH. To address these questions, we present NOAH, an in-situ **n**avigation interface for **o**verviewing and **a**nalyzing data **h**olistically. The goal of designing such an interface is to address the navigational challenges of spreadsheets. As a proof of concept, we have integrated NOAH as a data exploration plugin for DATASPREAD [dat], an open-source scalable web-based spreadsheet. The benefits of our proposed interface are as follows: NOAH *(a)* frees the user from moving back and forth between viewports by providing an *in-situ customizable overview* of data (see Figure 1a), *(b)* supports quick and easy navigation of data through panning and zooming interactions on the overview as an alternative to scrolling (see Figure 2a, 2b, 2c), *(c)* ensures interoperability among the overview and the spreadsheet so that users can relate the interactions on the overview with the raw data (see Figure 2d), and *(d)* creates and displays *contextual and historical information* automatically as users navigate the spreadsheet using the overview so that users don't lose context of their navigation (see Figure 1d and 1e).

Outline. The rest of the paper is organized as follows: In Section 2 we first present a usage scenario on a real world dataset and walk through the different navigational tasks related to spreadsheets while discussing the benefits of having NOAH as a plugin for spreadsheets. In Section 3, we discuss the related work in detail. We then explain the design of the NOAH interface in Section 4. Finally, in Section 6, we present the design and outcomes of a user study evaluating the effectiveness of NOAH-like features for spreadsheet navigation.

2. Usage Scenario

To illustrate the benefits of NOAH, let us consider a scenario where Rebecca, an investigative journalist, is exploring the Inside Airbnb dataset [webb], a dataset of all of the Airbnb listings across different US cities. She is specifically interested in investigating the long-standing accusation that many listings in Airbnb are illegally run as hotel businesses (while avoiding taxes) [acc]. Moreover, she wants to identify a few of these listings for further investigation. As she explores the data using a typical spreadsheet, she encounters a number of challenges involving navigation. We also briefly mention how she can use NOAH to address these challenges.

Overview: Suppose Rebecca wants to start with analyzing one of the big cities. However, she doesn't know what are the different cities in the dataset and how many listings are there in each city. By simply looking at the data in the spreadsheet, she cannot find the answer to any of these questions. As a workaround, she can use the filter menu to check all the unique city names. To get a better overview, she can create a pivot table of cities in a new spreadsheet to find out the total number of listings per city. She eventually finds out the larger cities in US by comparing the number of listings for each city. Rebecca now wants to navigate to the listing in a specific city, say Chicago. However, to access the listings in Chicago she can't use the pivot table and has to resort to either scrolling or filtering.

Using NOAH, Rebecca can first choose to explore the data by city and skim over the overview to identify the larger US cities (see Figure 1a). She can click on specific bins (similar to Figure 2d) to access and view the corresponding spreadsheet data.

Steering: Suppose Rebecca now wants to find out the listings in Chicago that are frequently rented—listings with availability ≥ 60 are considered as frequently rented listings [webb]. She needs to use the `COUNTIF` formula to get this statistics. However, she cannot use pivot table to get this information as pivot tables do not support the `COUNTIF` formula. Therefore, she has to resort to tediously steering the data to identify the row ranges corresponding to Chicago and then write down the appropriate `COUNTIF` formula in a cell within the spreadsheet.

Using NOAH, Rebecca can simply issue the formula on the overview which adds a formula column with the availability statistics for each bin (see Figure 1b). She can then zoom into the *Chicago-Denver* bin to find out the availability statistics for Chicago. She finds out that a relatively higher number of listings are rented more than 60 days a year.

Analysis: She now wants to see whether the same pattern emerges

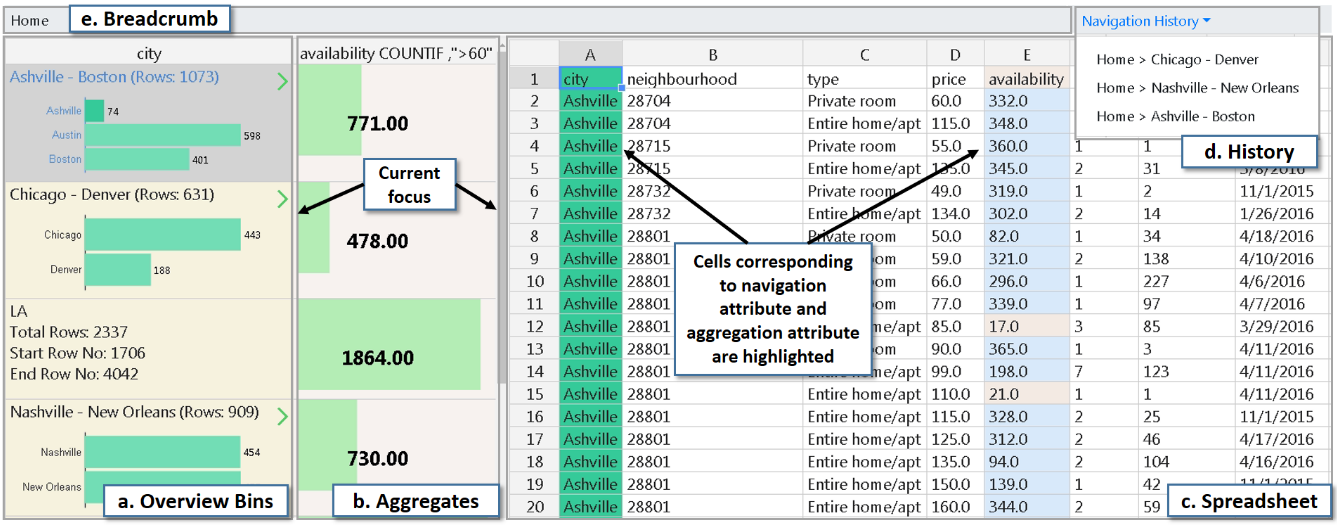


Figure 1: NOAH *interface: (a) and (b)zoomable overview with aggregate column, (c) spreadsheet, (d) navigation history displaying bins visited so far, and (e) breadcrumb showing the current navigation path.*

for other big cities, *e.g.*, New York City, Boston, and compare their availability statistics. To get the same results for those cities on a typical spreadsheet, she needs to repeat the same tedious process mentioned above while moving between different spreadsheet windows. The problem is further exacerbated if she wants to analyze the same statistics for all of the cities.

As Rebecca has already issued the formula on the NOAH overview, she does not need to reissue the formula to investigate similar statistics for other cities. She can simply navigate to different bins either through mouse clicks or zoom in operations. NOAH then computes and displays the statistics for the current view on demand (see Figure 3d).

Scrolling: To compare the availability statistics of the cities, she needs to access the COUNTIF formula results each city written in different cells within the spreadsheet. Therefore, she needs to again scroll up and down the spreadsheet to find those results. She could've used the named ranges option to manually store a reference to the cell where the formula was written. However, the process becomes extremely cumbersome as the number of cities being compared increases.

Using NOAH, Rebecca can find out all the bins she has explored so far by looking at the navigation history (see top right of Figure 1d). Her current context, *Chicago-Denver* in this case, is also highlighted in gray (see Figure 1a). She can also see the current navigation path on the top left (see Figure 1e).

Relate: Finally, Rebecca wants to quickly skim through the Chicago listings and identify a few listings with availability ≥ 60. Her goal is to use the information to pinpoint a few listings and investigate those listings in person. In Excel, she has to either issue an auto-filter to filter out all the values < 60 or use conditional formatting to highlight the listings ≥ 60. NOAH on the other hand, already highlights the listings with availability ≥ 60 in sky blue (see Figure 1c)and Rebecca can quickly spot those listings.

The aforementioned scenario highlights many of the navigational challenges spreadsheets users face. Throughout the rest of the paper, we refer to the same dataset and use the same tasks to explain the benefits of the features provided by NOAH for spreadsheet navigation. We now discuss the related work that addresses some of these problems.

3. Related Work

In this section, we present existing spreadsheet software and other alternatives that partially address the limitations of navigating spreadsheets explained in Section 1.

3.1. Existing Spreadsheet Software

We first discuss the features supported by both commercial spreadsheet software and other academic prototypes to address the navigational challenges.

3.1.1. Commercial Spreadsheet Software

There are several commercial spreadsheet software available in the market. Microsoft Excel is the most widely used spreadsheet tool and supports a suite of features that many of the other competing spreadsheets don't provide. On the other hand, Google Sheets is the most widely used and feature rich web-based spreadsheet. Airtable, a spreadsheet-database hybrid that allows users to store, organize, and collaborate on information online while a supporting a variety of data types and interactions. However, it does not address any of the navigation challenges of spreadsheets. Therefore, we discuss how the Excel and Google sheets provide support for navigation.

Microsoft Excel. In Section 1, we explained the named ranges and pivot table features of Microsoft Excel that provide support for scrolling and analysis, respectively. Excel also provides another alternative to pivot tables—SUBTOTAL [sub] which generates a limited set of summary statistics of the data and adds the summary as a row in the spreadsheet—each summary is added at the end of the corresponding subset of data. Initially SUBTOTAL only displays the summary. Users can collapse the summary to view the actual spreadsheet data. However, as the number of subsets within the data increases, the size of the summary itself can become very large, *e.g.*, for numeric data. This introduces visual discontinuity when a user compares the summary of different subsets.

Google Sheets. Google Sheets Explore [weba] provides users with statistics on the data by auto-generating charts, thus partially addressing the analysis challenges. Furthermore, users can specify queries to the system (similar to web search) asking different summary statistics about the data. However, the system doesn't address any of the navigational challenges related to scrolling and steering.

3.1.2. Academic Prototypes

Existing academic work primarily focuses on either enhancing spreadsheet capabilities by adding richer sets of operations to the spreadsheet or increasing the scalability of the spreadsheets.

Enhancing Spreadsheet Capabilities. *VisSh* [NB00] extends spreadsheets by replacing the formula oriented computation model with a declarative language that ingests a list of values instead of constants or formulae and generates any type of output, even charts, within cells. *Gneiss* [CM16] enables users to load hierarchical data into spreadsheets, order data along different hierarchical attributes, and perform hierarchical joins on the data. *Related Worksheets* [BKM11] provides users an alternative to databases when working with smaller datasets and enables users to join different sheets within a single workbook.

Scalable Spreadsheets. Smart drill down [JGMP15] can support arbitrarily large data and uses a rule-based approach to summarize spreadsheet data as a collection of rules. Users to drill down to a specific rule to investigate the detailed data. However, users cannot perform any analysis operation on the summary. ABC [R*99] enables dynamic reordering of arbitrarily large datasets and allows users to keep interacting with the spreadsheet as data is being reordered. DATASPREAD [B*18] is a web-based spreadsheet that unifies spreadsheets and databases with the goal of enhancing the usability of databases and can handle datasets with billions of rows. Neither ABC nor DATASPREAD address any of navigation challenges of typical spreadsheets that are further magnified by the increasing scale of data

3.2. Spreadsheet Alternatives

In this section, we discuss overview+detail and zooming interfaces, and multiple coordinated views that address the navigational challenges in different domains.

Cockburn et al. [CKB09b] provided a detailed survey of overview+detail and zooming interfaces. To improve navigation within large documents, overview+detail interfaces [CGA06, KBR10] allow users to interact on an overview as they explore the document. Zooming interfaces [NBM*06,PGB02] provide a multi-granularity overview of the data and support interactions like zoom in/out to navigate through the granularity. We also follow the same analogy of providing an overview of the spreadsheet first and then allowing the users to drill down further.

Multiple coordinated views, *e.g.*, *Snap* [NS00], *Elastic Documents* [BLE18] connect multiple views, for example, an overview and a detailed data view while enabling coordination between these views through brushing and linking. Similarly, NOAH connects tabular spreadsheets with an overview, and updates the state of the spreadsheet as users interact with the overview and vice-versa.

4. NOAH: A Zoomable Overview Interface

In this section, we present NOAH, a zoomable overview interface for tabular spreadsheet data. We first discuss our design considerations for the interface and then present the features and interactions that NOAH offers.

4.1. Why a Zoomable Overview?

We introduce an in-situ an overview interface for spreadsheets to address the navigation challenges related to scrolling, steering, and analysis. We iterated through several interface mockups to furnish the design of the overview interface while drawing inspiration from existing overview+detail, and zooming interfaces [CKB09b].

A standard approach for displaying an overview is to show a collection of thumbnails, similar to Microsoft PowerPoint, Adobe Reader, etc. However, displaying too many thumbnails in the overview results in increased scrolling to access distant thumbnails, thus increasing visual discontinuity. On the other hand, displaying too few thumbnails reduces visual discontinuity, but at the cost of visual clarity—the thumbnails appear cluttered and fail to represent the underlying data clearly [CKB09b]. As a result, there is a trade-off between visual discontinuity and visual clarity as we contemplate the design of an overview interface for spreadsheets.

To strike a balance between these two objectives, we adopt the visual information seeking mantra [Shn03] of providing a bird's eye view of the data first and then displaying details on demand as the user interacts with the overview. We design a zoomable overview interface which abstracts the data at varying granularities. We divide each granularity into non-overlapping groups of data called bins (*i.e.*, thumbnails of the overview). The highest level of the granularity provides a global view with less visual clarity and reduced visual discontinuity. We support different operations *e.g.*, clicking, semantic zooming [PF93] on the bins of the overview to support spreadsheet navigation. Moreover, users can leverage the overview to perform analytical operations on spreadsheet data. Finally, users can customize the structure of the overview generated by NOAH to accommodate their purpose of navigation. We now discuss the construction of the overview interface.

4.2. Overview Interface Construction

NOAH constructs the overview of the spreadsheet along with a single attribute at any given point in time—the attribute can be either textual or numeric. The multi-granularity overview provides a general context for understanding the spreadsheet data and helps to differentiate the bins. These bins act as landmarks on the overview interface and allow users to quickly get to the desired data by skipping over irrelevant bins and then zooming into the relevant ones. Each bin contains summary information regarding the spreadsheet data/region it spans, *e.g.*, the starting row and ending row index in the spreadsheet, the total number of rows the region spans. Additionally, a bin may also display a histogram of its children in the next level, *e.g.*, in Figure 2a, the top bin has three children in the next level. Therefore, it displays a histogram of its children.

At any granularity, the grouping of the bins is essentially a equi-depth histogram constructed on the attribute by which the data is organized, *e.g.*, city. Equi-depth histograms are commonly used in many applications, *e.g.*, query optimizers in database systems, for understanding the statistical properties of the underlying data and therefore, are suitable representations for our overview interface. All the bins of an equi-depth histogram contain the same number of items where each item corresponds to one value of the attribute on which the overview is constructed. For example, when constructing the overview on the city attribute of the Airbnb data, the city names are items within the bins. The bins are constructed top-down (see Figure 2d)—NOAH divides each of the bins at level k into new bins to construct the next lower level $k-1$, again, by applying the same concept of equi-depth histogram construction. If all the values of an attribute are unique, each bin of the equi-depth histogram will contain the same number of items, where each item corresponds to one unique value of the attribute. However, the same value may be repeated more than once. For example, in the Airbnb data, there are multiple listings that belong to the same city. Therefore, constructing an equi-depth histogram on the attribute city will result in consecutive bins sharing elements of the same unique value. In that case, we construct the best effort equi-depth histogram which is as close to an equi-depth histogram as possible. We explain the operations and interactions supported by NOAH next.

4.3. Features and Interactions

We now present different features of NOAH and the set of allowable interactions that enhances user's navigation experience with spreadsheets.

4.3.1. In-situ Overview.

As explained earlier, the overview interface exists in-situ with the spreadsheet. NOAH currently supports overview construction only on numeric and textual data. Date and timestamps are not supported by DATASPREAD, the web-spreadsheet hosting NOAH and therefore, excluded from consideration for now. Figure 1 shows an overview of the right-hand side spreadsheet. The displayed overview changes as users perform different navigation operations on the overview interface (Figure 2). As users may perform ad-hoc operations on the overview interface, NOAH constructs the overview on demand. Users interact with the overview to view and navigate the spreadsheet data.

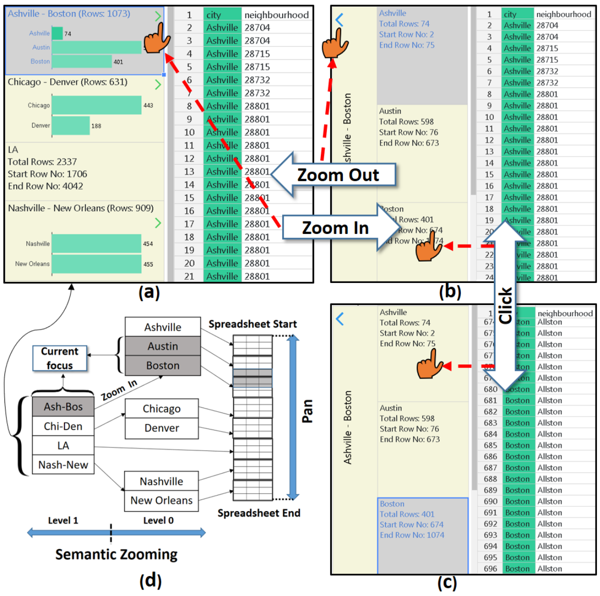


Figure 2: *Navigational operations. (a) The overview at the highest level of granularity, (b) a zoomed in view of the* Ashville-Boston *bin, (c) as user clicks on the* Boston *bin, the listings of Boston is displayed on the spreadsheet. The* Boston *bin is highlighted in gray to indicate users current focus, (d) conceptualizing the multi-granularity overview interface: clicking, semantic zooming, and context+focus on the binned overview.*

Navigational Operations. NOAH allows users to navigate the spreadsheet data by clicking on any of the bins. When a user clicks on a specific bin NOAH displays the corresponding spreadsheet data. As a result, users can jump to a specific location within the spreadsheet without having to scroll endlessly. For example, in Figure 2b, as the user clicks on the *Boston* bin, the data corresponding to Boston is displayed on the spreadsheet (Figure 2c). Moreover, users can zoom into a specific bin to view a more fine-grained overview of the underlying data or zoom out of a specific collection of bins to view a more coarse-grained overview. The zooming operation is essentially a semantic zooming [PF93]—as users zoom into a specific bin within the current granularity, we display a more detailed view at the next lower level granularity. For example, in Figure 2a, from the bin *Ashville-Boston* when the user zooms in to the next level, NOAH displays the bins *Ashville*, *Austin*, and *Boston* (Figure 2b). Several domains, *e.g.*, online maps, program visualization, adopt semantic zooming to ensure improved readability and comprehension of the information displayed [SGKC03]. If the user zooms out of the current granularity, again NOAH displays the bins *Ashville-Boston*, *Chicago-Denver*, *LA*, and *Nashville-New Orleans*. Any bin that contains multiple unique values can be zoomed in to. Users can continue to zoom into a bin until there's only a bin corresponding to just one unique value. For example, in Figure 2d, at level 0, each bin corresponds to one city in the spreadsheet. Therefore, users can only click on those bins, but cannot zoom into.

4.3.2. Customizing the Overview.

As NOAH constructs the overview interface automatically, the grouping may not fit the user's purpose of navigation. Therefore, NOAH allows users to customize the organization of the overview based on their own preferences. Users can merge multiple consecutive bins into a single bin or split a specific bin into multiple bins. The other operations include merging all the bins into one single bin and collapsing all the bins into singular bins, *i.e.*, one unique value per bin. For example, let's assume the user wants to compare some summary statistics of two cities, Boston and Chicago (see Figure 3). However, in the current organization of the overview of the two cities are located in two different bins (see Figure 3a). Using the bin customization feature, the user can merge the two bins *Ashville-Boston* and *Chicago-Denver* which creates a new bin *Ashville-Denver*. Users can now zoom into this bin and compare the summary statistics of the two cities in the same view resulting in a reduced visual discontinuity.

4.3.3. Leveraging the Overview for Analysis

To enable in-situ analysis of data, NOAH enables users to issue traditional spreadsheet formulae on the overview bins. This is equivalent to selecting a subset of data on the spreadsheet, *i.e.*, steering and then executing the same formula. Therefore, users can avoid steering the data when they are performing analysis operations using NOAH. The results of the analysis are displayed as a new column called *aggregate column* in the overview interface (see Figure 1b). Each entry in the column displays the formula results corresponding to a bin of the current granularity being displayed on the overview interface. The corresponding spreadsheet formula is displayed as a header of the aggregate column. For example, in Figure 3c, the aggregate column displays four aggregate statistics corresponding to the four different bins. Users can choose to perform different analysis operations on the overview simultaneously. As user issues a new analysis operation, a new aggregate column is added to the overview.

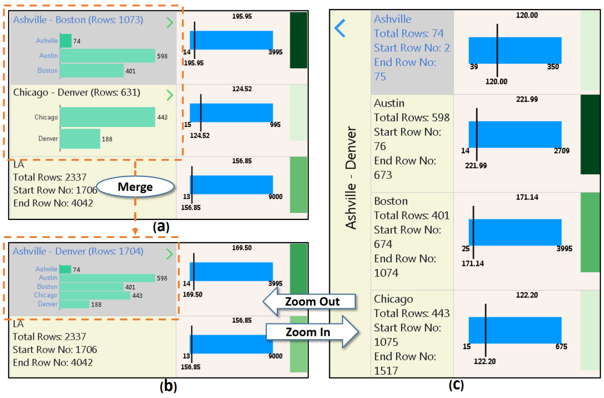


Figure 3: *Analysis and customization. (a) Chart view of the aggregate column. (b) A new bin created by merging the top two bins. This also triggers a re-computation of the aggregate column, and (c) zooming into the newly created bin, again triggers a recomputation of the aggregate column.*

To create an aggregate column, users are not required to explicitly type any formula, rather select appropriate formula from a drop-down menu, and provide necessary formula parameters, *e.g.*, input range, condition, to a web form. Currently, NOAH supports statistical and mathematical spreadsheet formulae that operate over a range of data, *i.e.*, takes a range parameter as input. We have further classified the supported formulae into five different categories: a) summary (*e.g.*, min, max, average), b) frequency (*e.g.*, mode, large, small), c) conditional (*e.g.*, countif, sumif), d) spread (*e.g.*, var,stdev), and e) others (*e.g.*, sum, count). Users can choose to view the result of a formula can be displayed either as raw value or charts and can toggle between views. Raw values are displayed along with a colored bar called *value bar* whose length is proportional to the corresponding aggregate value. This enables the users to visually compare the results of different bins. The chart representation varies depending on the formula type (see Figure 4).

As users zoom in and out of the overview interface, the results within an aggregate column continue to be updated as the granularity changes (see Figure 3). This is an added advantage of NOAH as users don't need to issue the same analysis operation over and over again as they zoom in and out. For example, in Figure 3b, the user can view the average price listing corresponding to different bins. To compare the results of Boston and Chicago, the user can zoom in, which automatically computes the average price of the bins in the next level (Figure 3c).

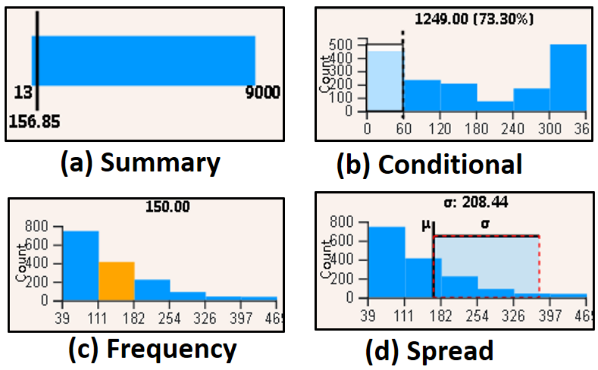


Figure 4: *Formula types and their chart representations.*

4.3.4. Contextual and Historical Information

As users operate on the spreadsheet, NOAH displays a user's current focus on the overview interface by highlighting the corresponding bin in gray. For example, in Figure 1, the *Chicago-Denver* bin is shown in gray, as the user is currently focused on the Chicago listings within the spreadsheet. Moreover, as the user pans across the spreadsheet the focus also moves along the overview (similar to the Sublime Text Editor). We explain the panning with an example later. NOAH also maintains the current navigation context through a breadcrumb-like feature which displays the current navigation path (see Figure 1e), similar to file directory browsing. NOAH also maintains a list of recent bins visited to allow users to move back and forth between the historical bins (see Figure 1d).

4.3.5. Interoperability

NOAH supports brushing and linking between the overview and the corresponding spreadsheet data. Clicking a bin is an example of brushing on the overview interface that populates the corresponding data on the spreadsheet. This is equivalent to a scroll operation in spreadsheets. On the other hand, as the user scrolls on the spreadsheet, the overview captures the current focus. For example, in Figure 2c, as the user clicks on the Boston bin, the spreadsheet displays the Boston listings. Conversely, as the user scrolls up, both Austin and Boston listings appear in the current window of the spreadsheet. Therefore, both the Austin and Boston bins are highlighted in the overview (see Figure 2d). Moreover, the user chooses to explore the spreadsheet by an attribute, the corresponding spreadsheet column is highlighted in lime green (see Figure 1) providing a visual cue. Similarly, when users issue an analysis operation on overview, the spreadsheet column corresponding to the resulting aggregate column is highlighted in grayish orange. For conditional formulae like COUNTIF, cells that satisfy the condition are highlighted in sky blue, *e.g.*, in Figure 1), the cells with availability ≥ 60 are colored in sky blue.

5. System Design

In this section, we provide an overview of the infrastructure of NOAH. We integrate NOAH as a data exploration plugin for DATASPREAD [dat], an open-source scalable web-based spreadsheet.

5.1. Underlying Data Structure(s)

The primary challenge in introducing the in-situ overview interface for spreadsheets is to seamlessly integrate both the interfaces: this integration should allow users to effortlessly perform interactions on both interfaces, enabling interactive navigation of the spreadsheet, brushing and linking between both interfaces, and rapid zoom in/out as users explore different granularities. As explained in Section 4, the underlying data structure representing the overview interface is an equi-depth histogram. NOAH constructs the histogram on demand based on the attribute chosen by the user—at the beginning only the highest granularity is constructed. As users perform ad-hoc interactions on the data different components of the interface are constructed dynamically. For example, as user zooms into a specific bin, NOAH again constructs an equi-depth histogram on the data corresponding to that bin on demand. The histogram is constructed in-memory.

To enable seamless integration of the navigation interface with the spreadsheet data, we leverage the hierarchical positional indexes used by DATASPREAD [B*18] to access the spreadsheet data. The index is essentially an order statistics tree [B*18] built on the position (*e.g.*, row number) of the spreadsheet data. For any given reordering of the data, the positional index is updated accordingly. Each bin of the histogram leverages the positional index to store the starting and ending spreadsheet position of each of its unique items. For example, the *Chicago-Denver* bin stores the starting and ending spreadsheet position of Chicago, and Denver. Therefore, NOAH can be integrated to any spreadsheet and requires only access to the positional mapping structure of that spreadsheet.

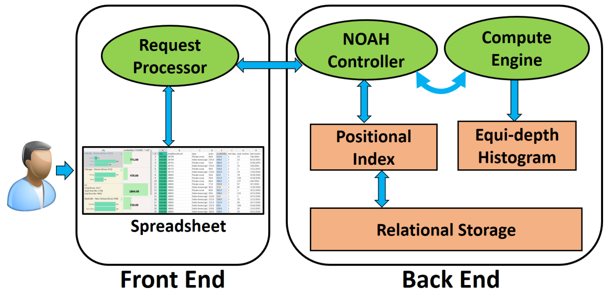


Figure 5: *System architecture.*

5.2. System Architecture

We now explain the system architecture of NOAH. The NOAH client is a web-based front-end that captures user input and renders both the overview interface and the spreadsheet based on the results returned by the back-end. The front-end is responsible for capturing user input and rendering components of the overview interface based on the response returned by the back-end. The web front-end for NOAH comprising four parts: (a) a zoomable overview; (b) an aggregate column depicting analysis results; (c) the actual spreadsheet, (d) navigation history, and (e) a breadcrumb showing current navigation path. NOAH supports ad-hoc interactions on the interface. The *request processor* receives the interactions from the user and issues a request to back end.

As the equi-depth histogram underlying NOAH's overview interface leverages the positional index of the spreadsheet the spreadsheet and the overview remain interoperable. The back end navigation controller receives the request from the front end. After processing the request corresponding to a user interaction, the *request processor* sends a response to the front-end encoded in *json*. For requests involving the spreadsheet, the *request processor* leverages the *positional index* to access the spreadsheet data. For requests involving the overview interface, *e.g.*, zoom in/out, *request processor* leverages the *compute engine* to manages the equi-depth histogram on demand. The compute engine is also responsible for processing the analytical operations. We leverage DATASPREAD's built in formula engine to support the analytical operations.

6. Evaluation Study Design

In this section, we present the design of a user study to evaluate whether the addition of NOAH-like features to spreadsheets assist in addressing the navigational challenges in spreadsheets related to scrolling, steering, and analysis.

6.1. Study Design and Participants

The study was designed to answer the following research questions:

RQ1 How does the addition of an overview to spreadsheets address the challenges of performing (a) scrolling, (b) steering, and (c) analysis tasks?

RQ2 How does the interoperability between spreadsheets and the overview help users in analyzing the spreadsheet data?

RQ3 How does bin customization accommodate a user's purpose of data exploration compared to the automated overview?

RQ4 How does a user's navigational experience vary with the addition of NOAH-like features to traditional spreadsheets?

Study Design. We conducted a 2×2 (2 datasets, 2 tools) mixed design within subject study. We explain the datasets next. The two tools used in the study were: Microsoft Excel, and NOAH (integrated with DATASPREAD). The study consisted of three phases: (a) an introductory phase explaining the essential features of NOAH using a video tutorial, followed by a warm up session where participants explored a flight dataset [webc] in NOAH to familiarize themselves with the features, (b) a quiz phase where the participants first used both the tools to perform targeted tasks on two different datasets and then completed one survey for each tool to provide feedback on their impressions about the tool, and (c) a semi-structured interview to collect qualitative data regarding the quiz phase.

Datasets. We used two datasets for the study—birdstrikes (used in Keshif [YEB18] and Voyager [WMA*]), and airbnb [webb]. These datasets were chosen for real-world interest to a general audience. The birdstrikes dataset has 10,868 records and 14 attributes (8 categorical, 1 spatial region, 1 temporal, 4 numeric). To ensure a fair comparison across tools, we created a sampled version of the original airbnb dataset with 10,925 records, by uniformly sampling 10% of the records from each US city. The dataset contains 15 attributes (6 categorical, 2 spatial region, 1 temporal, 6 numeric).

Participants. We recruited 20 participants (11 female, 9 male) via flyers across the university and via university email newsletter. The average age of the participants were 31.06 years ($\sigma = 12.44$). The participants came from different backgrounds, *e.g.*, engineering (seven), business (five), administration (five). During the recruitment process, prospective participants filled out an interest form where we asked them questions regarding their spreadsheet expertise, types of spreadsheet tasks performed, and usage of spreadsheet operations. We recruited participants having higher expertise (≥ 4 on a scale of one to five) with Microsoft Excel to ensure that participants' experience with spreadsheets didn't affect their performance during the quiz phase. Moreover, all of the participants were familiar with the basic mathematical and statistical functions supported by Excel. Each participant received \$10 per hour at the end of their session.

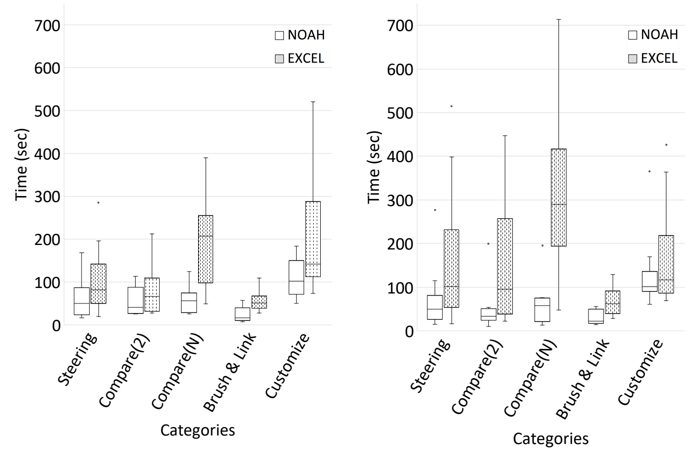
6.2. Study Procedure

We now explain each of the phases of the user study in detail.

Phase 1: Introduction to NOAH. We began the study by showing a six minute video tutorial explaining the features of NOAH on a dataset of all the flights across United States for January, 2018 [webc]. The participants then explored the same dataset using NOAH to familiarize themselves with the tool for about 10 minutes. The quiz phase began as soon as the participants finished their exploration.

Phase 2: The Quiz Phase. The purpose of the quiz phase was to evaluate the effectiveness NOAH-like features in addressing the spreadsheet limitations, *e.g.*, scrolling, steering, and analysis. Each participant performed specific tasks on the two datasets (birdstrikes and airbnb), using Excel for one dataset and NOAH for another. The order of the datasets were alternated between consecutive participants. The order of the tools were alternated between every two participants. We developed an online JavaScript based quiz system that stored users' responses as well as submission times for each tasks in a server. We also recorded the participants' interactions with both the tools using a screen capture software. During the quiz phase participants were allowed to refer to the web for any help regarding the Excel tasks. We also provided a printed handout to the participants that contained screenshots explaining the features of NOAH. We now explain the quiz tasks.

Quiz Tasks. To answer the first three research questions, we designed six tasks grouped into four categories. *i.e.*, steer (two tasks), compare (two tasks), brush and link (one task), and customize (one task). The questions were presented with increasing level of difficulty and simulated an independent exploration of a spreadsheet as demonstrated in Section 2. Participants began the quiz phase with a steer task to evaluate *RQ1a*, *e.g.*, calculating the total number of airbnb listings with availability ≥ 60 for a specific city. The steer task was followed by a brush and link task that was designed to evaluate *RQ2*, *e.g.*, looking up listings in a city that satisfied the condition of the steering task, *i.e.*, availability > 60. Then participants performed another steer task. Next, they performed two compare tasks Compare (2) and Compare (N), respectively, which evaluated both *RQ1a* and *RQ1c*. For the first task, Compare (2), participants had to navigate through the spreadsheet to compare the statistics of two subsets of data, *e.g.*, comparing the availability statistics of two cities, whereas the other task (Compare(N)) involved comparing all the subsets. The final task was the customize task which was designed to evaluate *RQ3*, *e.g.*, comparing average price of listings within an availability range.



(a) birdstrikes **(b) airbnb**

Figure 6: *Per category box plot of submission times for (a) birdstrikes, and (b) airbnb dataset.*

Survey. After completing the tasks for each tool, participants filled out a survey asking 15 likert-scale questions related to comprehensibility, level of satisfaction, ease of use, and speed of use of the corresponding tool. Users were also asked to mention the positive and negative aspects of both tools.

Evaluation. We evaluated both the accuracy (either 0 or 1) and completion time for each of the six tasks. We also gathered qualitative data regarding participants' interactions with the tools for each category of questions by analyzing the video recordings. Moreover, we aggregated the survey responses for each of the aforementioned criteria to quantify the usability of both the tools. **Phase 3: Interview Phase** Following the and survey, we conducted a semi-structured interview to identify the participants' preferred tools for different tasks and the reasoning behind their choices. We also asked participants to comment on the usefulness of different features provided by NOAH and Excel. We recorded the interviews for transcription purpose.

7. Results

In this section, we analyze the quantitative and qualitative data collected during the quiz, and the interview phases to address the research questions mentioned in Section 6.1. We first discuss the results of the quiz phase tasks. We then present participants' experiences in using both NOAH and Excel based on the interview and survey responses.

7.1. Analysis of the Quiz Phase

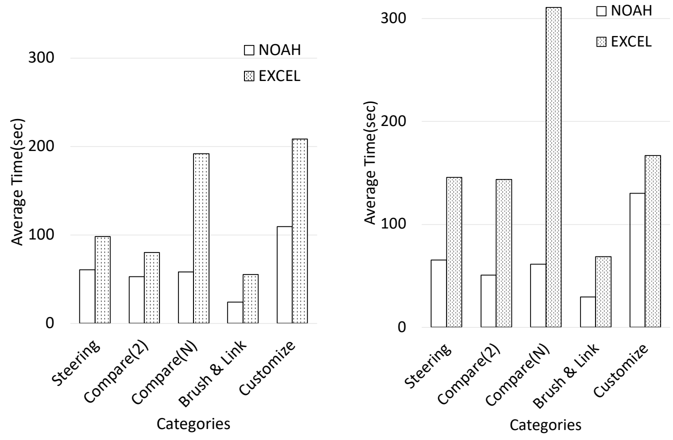
We now analyze the participants' task completion time and accuracy for NOAH with Excel.

7.1.1. Comparison of Task Completion Time

In Figure 7a and 7b, we show the average submission times of participants for the four categories of tasks, for the birdstrikes and airbnb dataset, respectively. The horizontal axis represents the categories whereas the vertical axis represents the average time of submission (in seconds).

Summary: We find: *(a)* For both the datasets, compared to Excel, participants took almost half the time in NOAH to complete the tasks for each category, and *(b)* Unlike Excel, irrespective of the dataset, participants' task completion times per task category were similar in NOAH.

Distribution of Completion Times. To get a clearer understanding of participants submission trends, in Figure 6a and 6b, we show the box plot analysis of participants' submission times for the four category of tasks, for both the datasets. The horizontal axis, represents the category of questions whereas the vertical axis represents the average time of submission (in seconds). For all of the tasks except the Compare (2) task in the birdstrikes dataset and the customize task in the airbnb dataset, *participants median submission time for* NOAH *was smaller than the fastest submission time for Excel*. This further confirms the efficiency of the feature extensions introduced by NOAH. To evaluate the statistical significance of the tasks completion times for all of the tasks, we ran *Mann-Whitney's U test* (as completion times did not follow a normal distribution). *For all of the tasks except the customize task, we found a significant effect of the tools, i.e., the response times for the tasks significantly differed by the choice of the tool* (see Table 1). We now discuss our observations for each task in detail.



(a) birdstrikes **(b) airbnb**

Figure 7: *Per category average submission time for (a) birdstrikes, and (b) airbnb dataset.*

Steer. When using Excel, participants took less time to submit steering tasks for the birdstrikes dataset compared to the airbnb dataset. The steer tasks for the birdstrikes dataset involved counting the number of 1s in a column with binary values. Analysis of the video recordings revealed that, many of the participants used the *autosum* feature to quickly obtain the counts—summation of binary values is equal to the number of 1s in the collection. Alternatively, some participants used the status bar at the bottom of the spreadsheet which displays the summation of all the cells in a selected column. In both cases, participants avoided steering the data. However, there were participants who first steered the data and then executed the formula. For the airbnb dataset, the column involved in the steer tasks had 365 different values. As a results, participants could not use the shortcuts mentioned above and ended up steering the data, resulting in higher task completion times. *In* NOAH*, participants avoided steering the data by issuing an aggregate formula on the overview interface. This indicates that avoidance of steering may improve spreadsheet task completion time*.

Compare. Each comparison task in Excel required users to steer the data and then compare the results. As explained earlier, may participants avoided steering for the birdstrikes dataset which resulted in faster completion for both the tasks compared to the airbnb dataset. For *Compare (N)* tasks in Excel, participants had to perform *N* comparisons, each time repeating similar tasks, *e.g.*, steering, autosum. As a result, the *Compare (N)* tasks had higher submission time than the *Compare (2)* tasks. However, the completion times were faster in NOAH than Excel irrespective of the comparison tasks and datasets. NOAH *offers three additional benefits apart from avoidance of steering that may have contributed to such improvement*: a) participants used the navigation features to access and compare different subsets of data, b) participants did not need to reissue the aggregation formula for any of the bins they navigated to, as the NOAH automatically applied the formula, and c) participants used the value bars presented along with the formula results to visually compare different subsets.

Brush and Link. For this task, participants had to skim through all the cells in the current window in Excel, resulting in higher completion times. Again, participants had to skim over binary values when counting in the birdstrikes dataset versus 365 different values when counting in the airbnb dataset, resulting in even higher completion times for the later. *In* NOAH*, participants quickly identified the cells in the spreadsheet that did not satisfy the condition of the steering task* based on the cell color ("grayish orange").

Customize. For the airbnb dataset in Excel, this task involved filtering out 26 values from the filter menu compared to filtering 451 values for the birdstrikes dataset. As a result, participants ended up steering a larger range of data for the birdstrkes dataset and understandably, took more time to submit their responses. The same task took much less time in NOAH compared to Excel. However, *this task took more time compared to the other tasks in* NOAH *as it required participants to restructure the overview before any calculation could be performed.*

Question	Category	Time (p value)	Accuracy (p value)
$Q1$	Steer	**0.0007** (*)	**0.0033** (*)
$Q2$	Brush and Link	$\mathbf{2.49 \times 10^{-5}}$ (*)	0.7475
$Q3$	Steer	**0.0043** (*)	**0.0202** (*)
$Q4$	Compare(2)	**0.0154** (*)	1
$Q5$	Compare(N)	$\mathbf{5.83 \times 10^{-6}}$ (*)	0.48
$Q6$	Customize	0.1207	0.0959

Table 1: *Statistical significance test results. (*) indicates statistically significant.*

7.1.2. Comparison of Accuracy

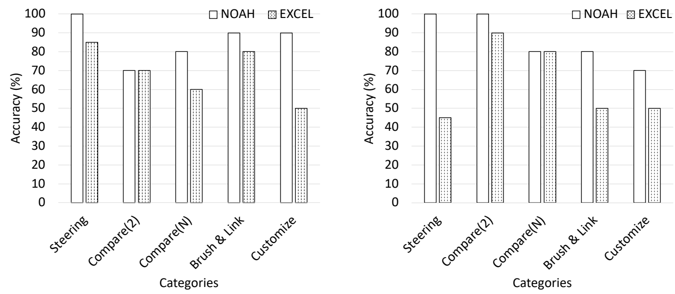
In Figure 8a and 8b, we show percentage of accurate submissions for the four category of tasks, for the birdstrikes and airbnb dataset, respectively. The horizontal axis represents the category of questions whereas the vertical axis represents the average accuracy.

Summary: For all of the tasks except for the Compare (2) task—for which the accuracy was same for both the tools—participants attained slightly higher accuracy with NOAH compared to Excel.

Accuracy Distribution. The accuracy statistics exhibit significant deviation for the steer tasks only. We ran the *Fisher's exact test* for statistical significance of categorical data that confirms the observation—*for steer tasks the percentage of the accuracy of submissions significantly differed by the choice of the tool.* We now discuss our observations for each task in detail.

Steer. For these tasks, participants had to manually select the ranges in Excel which led to incorrect selection ($N = 14$) and resulted in incorrect responses. The errors mostly occurred for the airbnb dataset ($N = 11$) where participants were not able to avoid steering unlike the birdstrikes data.

Compare. For compare (2) tasks on the birdstrikes dataset participants exhibited the same accuracy for both NOAH and Excel—a deviation of the trend. Analysis of the screen capture that revealed that between Florida—the correct choice—and California, three participants ($P12$, $P13$, and $P20$) chose the later when using NOAH. For *Compare(N)* tasks, the number of subsets participants had to compare were higher in the birdstikes dataset (50) compared to the airbnb dataset (16). As a result, participants exhibited lower accuracy for the birdstrikes dataset when using Excel. In NOAH, all of the participants first split all the bins to create N bins each corresponding to one state. Then participants panned across the overview to find out the desired bin as they compared the values. Again comparison between multiple values resulted in several incorrect submissions.



(a) **(b)**

Figure 8: *Per category average accuracy for (a) birdstrikes, and (b) airbnb dataset.*

Brush and Link. For this task, participants attained higher accuracy in NOAH by identifying relevant cells based on color. Whereas in Excel, participants had to manually skim through the column.

Customize. For this task again majority of the errors in Excel resulted from users failing to select the appropriate range of data. In NOAH for the airbnb dataset, two participants ($P11$ and $P18$) did not merge the bins while one participant aggregated an incorrect column, leading to lower accuracy compared to the birdstrikes dataset.

7.2. Interview Phase

We now present a qualitative analysis of the participants' perceptions of both tools based on their interview responses.

7.2.1. RQ1: Overview and Navigation

Without an overview of the data, participants found it difficult to perform different tasks on Excel—*I think Excel can get overwhelming if you have a lot of data in it and sometimes with that data finding things can be difficulty* ($P11$). The binned representation of the data in NOAH, on the other hand, helped users to comprehend the overall structure of the data better and prioritize the bin they want to visit next, thereby enabling faster navigation. Out of 20 participants 15 participants preferred the overview interface of NOAH for performing tasks involving navigation. For example, one participant ($P5$) commented: *I think it was just a little bit easier to navigate and find where things were because you could already see what bins had what.* Another participant ($P1$) commented: *I like*

NOAH *a lot better. It was a lot easier to look up different data and it was a lot quicker too*.

Moreover, participants found scrolling and steering to be cumbersome while issuing formula and performing comparisons of subsets of data— *The one thing with Excel, is I always try to go to the bottom of the data and type in the formula, and with something really long like this, the scrolling is a little bit cumbersome (P4).* Using NOAH, on the other hand, users can (a) avoid scrolling by using clicking or zooming operations, and (b) avoid steering by performing aggregate operations on the overview. These operations further automate the navigation process in spreadsheets, thereby simplifying data navigation and analysis. For example, one participant (*P*3) commented: *And that creates convenience sort of because then you don't have to memorize anything and using the system becomes easier.* Another participant commented: *With Noah you don't have to highlight every number ... versus excel you actually have to select everything (P12).*

However, the interactions on NOAH's overview interface are quite different from the interactions in the spreadsheet domain. This unfamiliarity led to some participants ($N = 5$) preferring Excel over NOAH. One participant (*P*11) commented: *Since I'm not used to spreadsheet data being presented that way, it took a little bit of getting used to.* Moreover, participants ($N = 2$) preferred having more control over the operations they were performing in Excel (*e.g.*, scrolling, steering) as opposed to the automations provided by NOAH—*I like the way that in Excel I can go look at my formula again if I wanted to make sure I'm doing everything I wanted to do correct (P6).*

7.2.2. RQ2: Brush and Link

Out of 20 participants 16 participants preferred the data highlighting feature of NOAH. Again participants preferred the automation provided by NOAH. Moreover, the colored cells acted as visual cues to the participants and focused their attention to relevant regions within the spreadsheet—*You didn't have to do any additional steps and it was a visual cue right there, made it very quick to count it up (P17).* One participant (*P*8) did not like the color codings in NOAH—*I thought the color choice wasn't the greatest.*

7.2.3. RQ3: Bin Customization

The bin customization feature enables users to personalize the overview based on the purpose of their navigation and analysis. One participant (*P*16) commented: *I did like the fact that it lets you take a data sheet and in some way containerize the stuff you care and the stuff you don't care about.* Out of 20 participants, 14 participants found the bin customization feature to be useful. In particular, participants preferred the feature compared to Excel's filtering feature when working with numeric data—*That was so much easier in* NOAH *than it was in Excel to be able to specify the range that you wanted it to go in (P17).* However, six participants found the feature to be difficult to use when working with textual data and would've preferred a pivot table-like single level overview—*I would have expected them to completely be split, and then I can merge them if I want to (P13).*

7.3. Survey Results

Table 2 shows the average rating for different metric, *e.g.*, ease of learning, both speed and ease of use, confidence, comprehensibility, satisfaction. All the ratings were on a scale of one to seven with one being the lowest and seven being the highest. The distribution of the ratings for none of the the criteria followed a normal distribution. Therefore, we performed a *Wilcoxon Signed-rank* test on the survey responses which shows that for all the criteria, the ratings significantly differed by the choice of the tool, *i.e.*, NOAH or Excel. The results in table 2 show that users preferred NOAH more than Excel in all the criteria. Notably, participants found NOAH to be easier to use compared to Excel while being faster in execution. Even though the features offered by NOAH were new to the participants they found it fairly easy to learn.

Metric	NOAH	Excel	p value
Ease of Learning	$\mu = 5.75$, $\sigma = 1.02$	$\mu = 4.22$, $\sigma = 1.41$	$\mathbf{1.49 \times 10^{-7}}$ (*)
Speed of Use	$\mu = 6.03$, $\sigma = 0.99$	$\mu = 4.22$, $\sigma = 1.65$	$\mathbf{1.68 \times 10^{-7}}$ (*)
Ease of Use	$\mu = 5.88$, $\sigma = 0.90$	$\mu = 4.33$, $\sigma = 1.71$	$\mathbf{7.85 \times 10^{-6}}$ (*)
Confidence	$\mu = 5.50$, $\sigma = 1.79$	$\mu = 4.60$, $\sigma = 1.50$	**0.0096** (*)
Comprehensibility	$\mu = 5.60$, $\sigma = 1.27$	$\mu = 4.48$, $\sigma = 1.65$	**0.0006** (*)
Satisfaction	$\mu = 5.48$, $\sigma = 1.16$	$\mu = 4.52$, $\sigma = 1.49$	**0.0018** (*)

Table 2: *Survey results. (*) indicates statistical significance.*

7.3.1. Limitations and Future Work.

The interviews covered majority of the advantages and disadvantages of both the systems. However, participants mentioned additional limitations of NOAH during the survey. By addressing these limitations, we can further improve the interface. Participants found some of the terminologies used in the interface—*e.g.*, explore, bin—to be quite unfamiliar ($N = 14$) for an Excel audience. Moreover, two participants didn't understand how the bins were constructed. These issues can be addressed by using more relatable terminologies and proper documentation. Two participants found the aggregations operations applied on the bins to be opaque compared to Excel where a user can directly manipulate the formula. In the next version, we can further display the appropriate formula for each bin as users hover over the corresponding cell on the aggregate column. One of participant (*P*5) drew attention to the fact that the color choices would make NOAH unusable for the colorblind population. We plan to address this issue in the next phase of our implementation. Finally, participants ($N = 5$) also noted the fact that NOAH currently does not support user defined formula, again a feature which is easy to implement.

8. Conclusions

In this paper, we present NOAH, a multi-granularity overview interface that allows users to navigate spreadsheet data in-situ as well as analyze the data using traditional spreadsheet formula. We integrated NOAH with an existing web-based spreadsheet and show how these extensions enhance a user's navigation experience with spreadsheets. We conducted a user study which demonstrated that participants found NOAH to be easy to use compared to Excel when navigating and analyzing large datasets in spreadsheets. The study showed that for different tasks involving navigation particpants were able to attain higher accuracy when using NOAH compared to spreadsheets while being quicker ($\approx 2X$ faster).

References

[acc] A Lone Data Whiz Is Fighting Airbnb—and Winning. https://www.wired.com/2017/02/a-lone-data-whiz-is-fighting-airbnb-and-winning/. 2

[B*18] BENDRE M., ET AL.: Towards a holistic integration of spreadsheets with databases: A scalable storage engine for presentational data management. In *ICDE* (April 2018). 4, 7

[BAM14] BRADBARD D. A., ALVIS C., MORRIS R.: Spreadsheet usage by management accountants: An exploratory study. *Journal of Accounting Education 32*, 4 (2014), 24–30. 1

[BI90] BEARD D. V., II J. Q. W.: Navigational techniques to improve the display of large two-dimensional spaces. *Behaviour & Information Technology 9*, 6 (1990), 451–466. 1

[BKM11] BAKKE E., KARGER D., MILLER R.: A spreadsheet-based user interface for managing plural relationships in structured data. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), ACM, pp. 2541–2550. 4

[BLE18] BADAM S. K., LIU Z., ELMQVIST N.: Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE transactions on visualization and computer graphics* (2018). 4

[CGA06] COCKBURN A., GUTWIN C., ALEXANDER J.: Faster document navigation with space-filling thumbnails. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 1–10. 4

[CKB09a] COCKBURN A., KARLSON A., BEDERSON B. B.: A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR) 41*, 1 (2009), 2. 1

[CKB09b] COCKBURN A., KARLSON A., BEDERSON B. B.: A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR) 41*, 1 (2009), 2. 4

[CM16] CHANG K. S.-P., MYERS B. A.: Using and exploring hierarchical data in spreadsheets. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 2497–2507. 4

[dat] DataSpread. https://dataspread.github.io/. 2, 7

[exc] How finance leaders can drive performance. https://enterprise.microsoft.com/en-gb/articles/roles/finance-leader/how-finance-leaders-can-drive-performance/. 1

[JGMP15] JOGLEKAR M., GARCIA-MOLINA H., PARAMESWARAN A.: Smart drill-down: A new data exploration operator. *Proceedings of the VLDB Endowment 8*, 12 (2015), 1928–1931. 4

[KBR10] KRATZ S., BRODIEN I., ROHS M.: Semi-automatic zooming for mobile map navigation. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (2010), ACM, pp. 63–72. 4

[KMB03] KRUCK S., MAHER J. J., BARKHI R.: Framework for cognitive skill acquisition and spreadsheet training. *Journal of Organizational and End User Computing (JOEUC) 15*, 1 (2003), 20–37. 1

[LBPFJ09] LAWSON B. R., BAKER K. R., POWELL S. G., FOSTER-JOHNSON L.: A comparison of spreadsheet users with different levels of experience. *Omega 37*, 3 (2009), 579–590. 1

[M*17] MORITZ D., ET AL.: Trust, but verify: Optimistic visualizations of approximate queries for exploring big data. In *CHI* (2017). URL: http://idl.cs.washington.edu/papers/trust-but-verify. 2

[nam] Define and use names in formulas. https://www.excel-easy.com/examples/names-in-formulas.html. 2

[NB00] NUNEZ F., BLAKE E.: Vissh: A data visualisation spreadsheet. In *Data Visualization 2000*. Springer, 2000, pp. 209–218. 4

[NBM*06] NEKRASOVSKI D., BODNAR A., MCGRENERE J., GUIMBRETIÈRE F., MUNZNER T.: An evaluation of pan & zoom and rubber sheet navigation with and without an overview. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 11–20. 4

[NM90] NARDI B. A., MILLER J. R.: *The spreadsheet interface: A basis for end user programming*. Hewlett-Packard Laboratories, 1990. 1, 2

[NS00] NORTH C., SHNEIDERMAN B.: Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies 53*, 5 (2000), 715–739. 4

[PF93] PERLIN K., FOX D.: Pad: an alternative approach to the computer interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (1993), ACM, pp. 57–64. 4, 5

[PGB02] PLAISANT C., GROSJEAN J., BEDERSON B. B.: Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.* (2002), IEEE, pp. 57–64. 4

[R*99] RAMAN V., ET AL.: Scalable spreadsheets for interactive data analysis. In *ACM SIGMOD Workshop on DMKD* (1999). 4

[SGKC03] SUMMERS K. L., GOLDSMITH T. E., KUBICA S., CAUDELL T. P.: An experimental evaluation of continuous semantic zooming in program. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on* (2003), IEEE, pp. 155–162. 5

[Shn03] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*. Elsevier, 2003, pp. 364–371. 4

[sub] SUBTOTAL. https://exceljet.net/excel-functions/excel-subtotal-function. 3

[weba] Google explore. https://www.techrepublic.com/article/how-to-use-the-new-google-explore-feature/. 4

[webb] The inside airbnb dataset. http://insideairbnb.com/get-the-data.html. 2, 8

[webc] The us flight dataset. https://www.transtats.bts.gov/. 8

[WMA*] WONGSUPHASAWAT K., MORITZ D., ANAND A., MACKINLAY J., HOWE B., HEER J.: Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization & Computer Graphics*, 1, 1–1. 8

[WPW99] WATTS-PEROTTI J., WOODS D. D.: How experienced users avoid getting lost in large display networks. *International Journal of Human-Computer Interaction 11*, 4 (1999), 269–299. 1, 2

[YEB18] YALÇIN M. A., ELMQVIST N., BEDERSON B. B.: Keshif: Rapid and expressive tabular data exploration for novices. *IEEE transactions on visualization and computer graphics 24*, 8 (2018), 2339–2352. 8