UF Herbert Wertheim College of Engineering
UNIVERSITY of FLORIDA

Department of Computer Science

Visual Summaries for Log Data

Under the guidance of Dr. Eric Ragan

A PRESENTATION BY

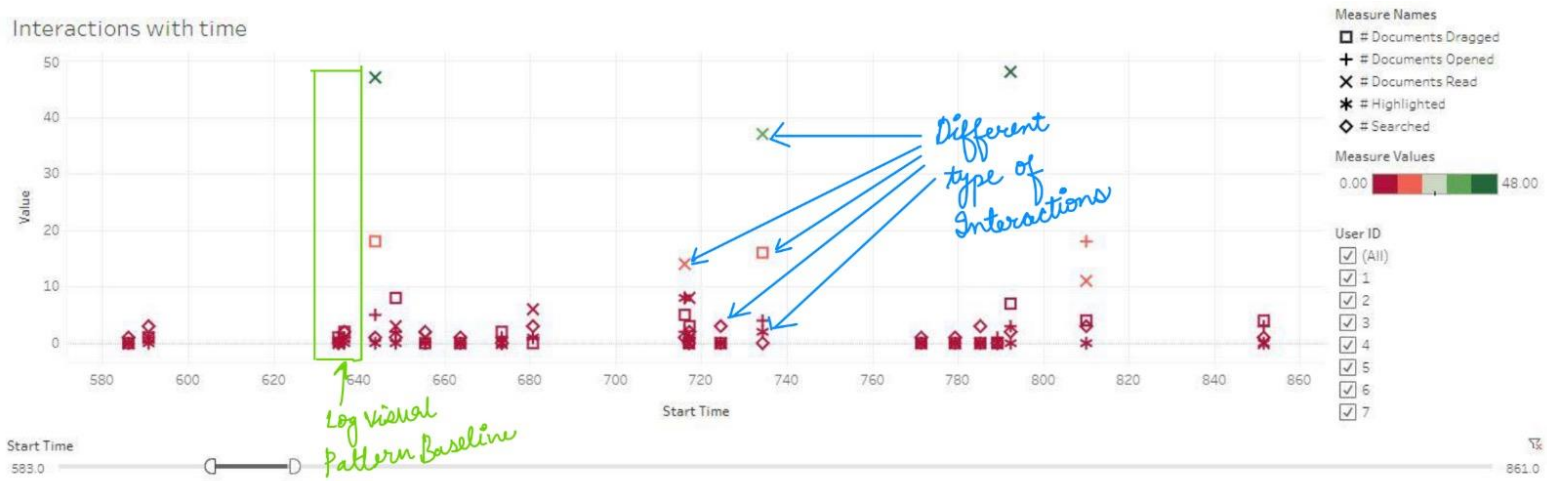
ANURAG KOLHE **ASEEM BARANWAL** **DEBIPRASAD DAS** **SRIJA MYANA**

Motivation

- Log data can give interesting revelations in identifying **trends and vulnerabilities.**
- Intelligence agencies can unearth any suspicious user behavior from these log files for inspecting any criminal activities.
- While incorporating different visualizations provided in our tool, a criminal agency can follow up on the analyst's behavior.
- In this project, we aim to visualize how an analyst has interacted with the system during a particular session.

Introduction

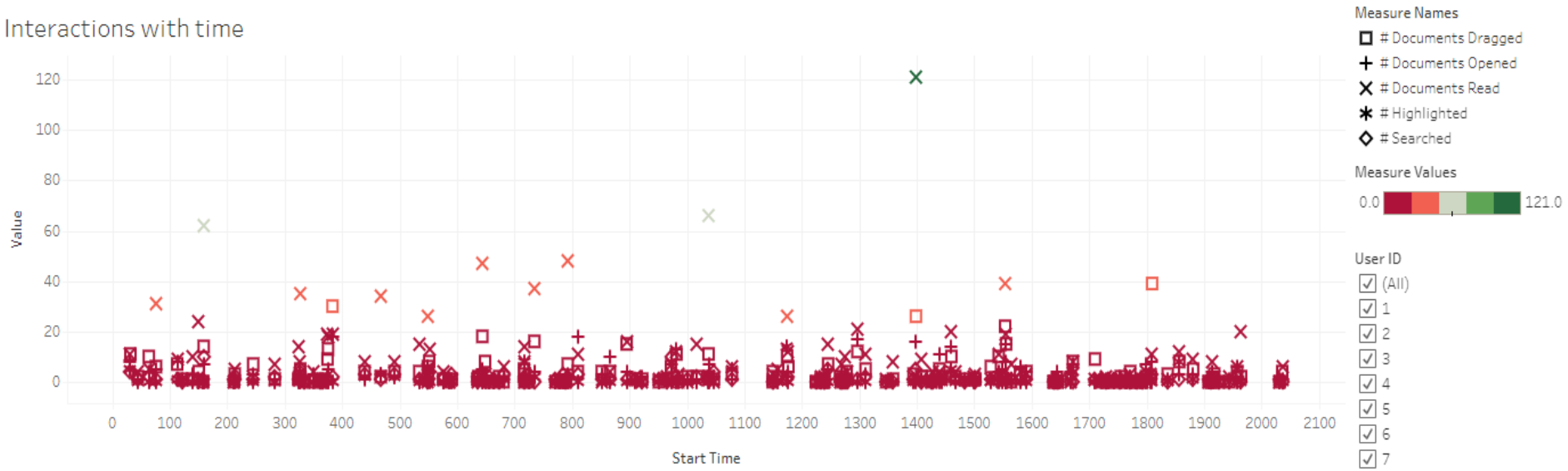
- Log Data is huge and finding interesting observations can be time-consuming and resource-intensive.
- Human brains are outstanding at understanding visual patterns and have a great rate of recall.
- Using different visual encodings, inferences can be drawn quickly without wasting valuable resources.



User ID	Most Read Document1	Searched Words1	Highlighted Words1
1	Armsdealing 65	['nigeria', 'gun']	['February 1, 2009']
	Armsdealing 84	['gun', 'nigeria']	[]
	Armsdealing 100	['gun', 'nigeria', 'nigeria gun']	[]
		['nigeria + gun', 'nigeria']	[]
	NA	['gun']	[]
		['kenya']	[]
		[]	[]
2	Armsdealing 81	[]	[' THE ENTIRE FUNDS TRANSFER WOULD STILL BE IN ITS ..

Dataset

- Three Multivariate Log Datasets with features along -
  - Users
  - Types of Interactions
  - Time

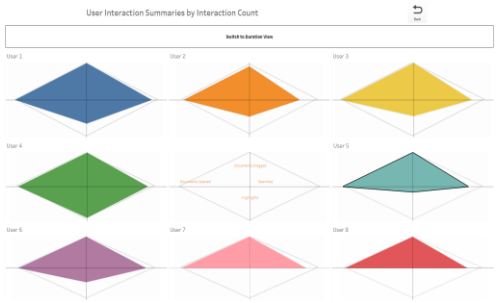


- Structure of Dataset
  - Document file containing contents of different documents read by Analysts.
  - User Interactions with time stamps and types of Interaction.
  - Segmentation of data into intervals for better sensemaking.

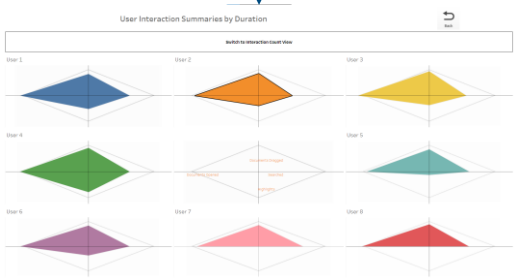
Analysis Task

- View the user interaction history.
- Guides in tracking user activities.
- Duration and Count of interactions using NLP techniques including stemming and lemmatization.
- Provides support in answering questions like -
  - Where an analyst spent most of the time and what activities were performed?
  - When did an analyst spend the most time reading the documents?
  - Interaction trends from beginning to end?
  - What is the **most read document** and **most searched keywords** in an analysis task?

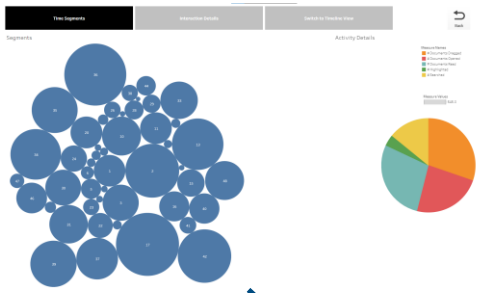
Visualization Structure



COUNT OVERVIEW

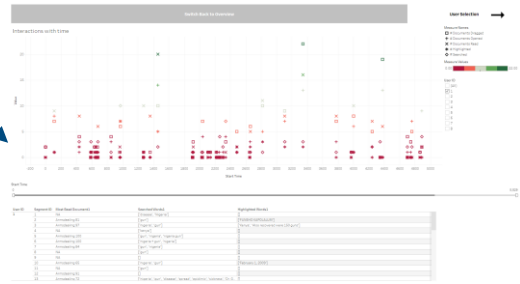


DURATION OVERVIEW



TIME SEGMENTS

INTERACTION VIEW



TIMELINE VIEW

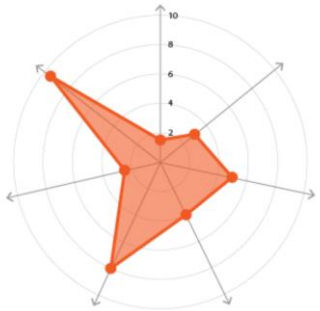
Visualization

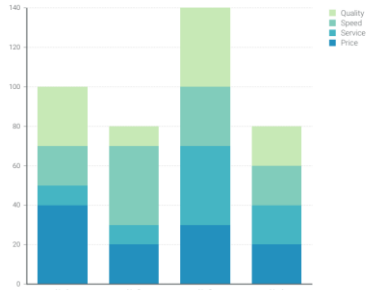
- Visual Encodings Used
  - Position
  - Length
  - Angle/Slope
  - Area
  - Color

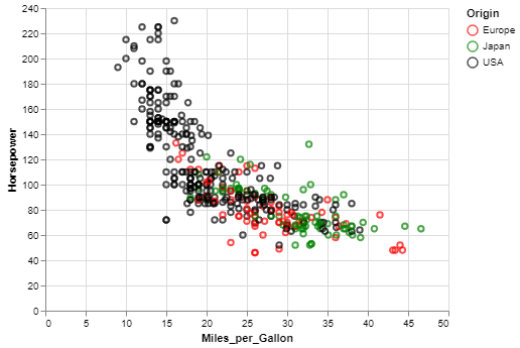
- Principles Used
  - Shneiderman's Mantra
  - Bertin's Visual Attributes
  - Cleveland and McGill Attribute Ranking

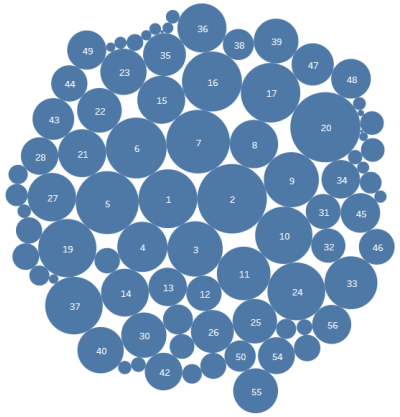
Visualization

- Visual Representations Used
  - Star Plots
  - Stacked Bar Chart
  - Scatter Plot
  - Bubble Chart
  - Pie Chart









Evaluation

- What's good
  - Concise summarization of how a user is interacting using different visualizations.
  - Provides an overview of how a user is spending his or her time.
  - Presents the change in the user's behavioral patterns with time in a chronological fashion.

- What's not good
  - Lacks dynamic loading of the dataset.
  - Cluttered visual representation when multiple time segments are present.
  - Absence of dynamic creation of Star Plots.

- Evaluation methodologies
  - Conducted peer-to-peer reviews from 12 people from different educational backgrounds.
  - Individual feedback was incorporated during interim demos.

Limitations and Contributions

- Limitations
  - No feature to load the dataset dynamically in UI.
  - Slow visual performance over larger data because of Tableau.
  - No functionality is provided for comparing multiple users' actions.

- Contributions
  - Our tool aids analysts in discovering interesting findings in a massive user log dataset.
  - A person's motives can be traced throughout time.
  - Investigating firms can exploit these features to defend some criminal accusations.

Thank you!