To study the application of Data Visualization and Analysis tools

Mrs. Shibani Kulkarni,
Department of Computer Science,
 Dr. D. Y. Patil ACS College,
Pimpri, Pune-18

Ms. Neeta Takawale,
Department of Computer Science,
Dr. D. Y. Patil ACS College,
Pimpri, Pune-18

Abstract:
       In today's world the amount of data has been exploding. Companies capture trillions of bytes of information every day about their customers, suppliers, and operations. Physical devices such as mobile phones, smart phones, and various other communicating devices generate loads of data similarly data on social networking sites will continue to grow.
       Large pools of data that can be captured, communicated, aggregated, stored, and analyzed—is now part of every sector and function of the global economy. Hence it's high time to preserve and secure this data and apply tools that can better visualize and analyse data.
There are many data visualizing and analytical tools available in the market today and analyst are still working to improve these tools. In this paper we present a brief review of most of the data visualization and analysis tools. We have focused on two such tools i.e. "Data Wrangler" and "Tableau Public". Wrangler combines direct manipulation of visualized data, enabling analysts to iteratively explore the space of applicable operations and preview their effects. Through the case study presented in the paper we have shown that Wrangler significantly reduces editing time as compared to manual editing and one can focus and spend more time on analysis instead of editing." Tableau Public" turns data into any number of visualizations from simple to complex. It also provides drag and drop options into work area which we have applied on the data and presented in this paper.
Keywords: Data Wrangler, Tableau Public, Data cleaning, transformation, visualization.

Introduction:

There are significant advancements in technologies for data management and analysis, but it still remains time-consuming to inspect a data set and convert it into a form that allows meaningful analysis. To improve data quality, analysts must identify and solve issues such as misspellings, missing data, duplicate data, and outliers. Data Wrangler is an interactive system for creating data transformations with reduction in editing time.

Wrangler provides an inference engine that generates and orders suggested transforms in response to direct manipulation of a data table. Analysts can navigate the space of transforms by directly selecting data, indicating a desired transform via menus, or by modifying a related transform; each of these actions leads wrangler to further refine the set of suggestions. To help analysts understand the effects of an operation before they commit to it, Wrangler's natural language transform descriptions are augmented by novel transform previews that visualize transform results. Using Tableau Public tool we have generated a bar chart.

In this paper we have illustrated the use of two tools - "Data Wrangler" and "Tableau Public" by applying them on real data. In the first case study we have applied the Data Wrangler tool on "Workshop data " and result analysis of B.Sc. Computer Science students of Dr. D. Y. Patil ACS College. In the second case study we have applied Tableau tool to convert the result analysis data in graphical form.

Review of Data Visualization and Analysis tools :

Following are some of the best Visualization and analysis tools available today:

- **CSV Kit tool :**
       This tool is available from Christopher Gyroscope's GitHub account. It

has a host of Unix-like command-line tools for importing, analyzing and reformatting comma-separated data files. CSV Kit makes it quick and easy to preview, slice and summarize.
Skill Level: Expert.

- **R project:**

IT is a statistical analysis tool. This tool uses two simple commands that generate a data structure that can, in turn, be used by several SQL database formats. The SQL code will create a table, inferring the proper data type for each field as well as the insert commands for adding data to the table. It has a UNIX like interface that is familiar to most users and makes it.
Runs on: Any Windows, Mac or Linux system with Python installed.
Skill level: Expert

- **Data tables :**

This is a popular jQuery plug-in which was designed and created by Allan Jardine. It creates sortable, searchable HTML tables from a variety of data sources -- say, an existing, static HTML table, a JavaScript array, JSON or server-side SQL. In addition to sortable tables, results can be searched in real time (results are narrowed further with each search-entry keystroke).
Runs on **:** JavaScript-enabled Web browsers.
Skill Level : Expert.

- **Free Dive**

This is alpha project from the Knight Digital Media Center at UC Berkeley. This tool turns a Google Docs spreadsheet into an interactive, sortable database that can be posted on the Web.
In addition to text searching, one can include numerical range-based sliders. End users can easily create their own databases from spreadsheets without writing code. Free Dive's chief current attraction is the ability to create databases without programming; however, free Dive source code will be posted and available for use once the project is more mature.
Skill level : Advanced beginner.
Runs on : Current Web browsers

- **High charts JS** :

This is a JavaScript library from High soft Solutions that provides an easy way to create professional-looking interactive charts for the Web. JQuery, Mootools or Prototype is required.
There are many different chart types available, from basic line, bar, column and area charts to zoomable time series; each comes with six stylesheet options. Little customization is needed to get a sleek-looking chart -- and charts will display on IOS and Android devices as well as on desktop browsers.
Skill level : Intermediate to Expert.
Runs on : Web browsers

- **Mr. Data Converter :**

New York Times interactive graphics editor Shan Carter coded a tool that converts comma- or tab-delimited data into nine different formats. It's available as either a service on the Web or an open source. Mr. Data Converter can generate XML, JSON, ASP/VBScript or basic HTML table formatting as well as arrays in PHP, Python (as a dictionary) and Ruby. It will even generate MySQL code to create a table (guessing at field formats based on the data) and insert your data. If your data is in an Excel spreadsheet, you don't need to save it as a CSV or TSV; you can just copy and paste it into the tool.
Skill level : Beginner
Runs on : JavaScript-enabled Web browsers.

- **Panda Project :**

Panda was specifically designed for newsrooms, but could be used by any organization where individuals collect information on their desktops that would be worth sharing. Panda makes it simple to give others access to information that's been sitting on individuals' hard drives in different stand-alone spreadsheets.
Runs on **:** Must be hosted on Amazon EC2 or a server running Ubuntu 11.10. Clients can use any Web browser.

- **PowerPivot :**

This free plugin from Microsoft allows Excel 2010 to handle massively large data sets much more efficiently than the basic version of Excel. It also lets Excel act like a

relational database by adding the capacity to truly join columns in different tables instead of relying on Excel. PowerPivot can handle million*s* of records -- data sets. And by joining tables, one can make more "intelligent" pivot tables and charts to explore and visualize large data sets with Excel's point-and-click interface.

Skill level : Intermediate

Runs on : Excel 2010 on Windows only.

- **Weave :**

This is a general-purpose visualization platform allows creation of interactive dashboards with multiple, related visualizations -- for example, a bar chart, scatter plot and map. The open-source project was created by the University of Massachusetts at Lowell in partnership with a consortium of government agencies and is still in beta. The visualizations are slick and highly interactive; clicking an area in one visualization also affects others in the dashboard. The platform includes powerful statistical analysis capabilities. Users can create their own visualizations on a Weave-based Web system, or save and alter the tools and appearances of visualizations that have been publicly shared by others.

Skill level**:** Intermediate for those just creating visualizations; Expert for those implementing a

Weave system.

Runs on**:** Flash-enabled browsers. Server requires a Java servlet container (Tomcat or Glassfish,

MySQL or PostgreSQL, Linux and Adobe Flex 3.6 SDK).

Though there are many tools available in the market we choose "Data Wrangler " and "Tableau Public" as our requirement was to convert unstructured data into tabular form. Also we wanted a tool that could easily convert the result analysis of students into graphical form.

Data Wrangler:

Data wrangler provides easy transformations through various options:

Wrangler transformation language options are as follows:

- Map transforms map one input data row to zero, one, or multiple output rows.
- Lookups and joins incorporate data from external tables.
- Wrangler includes extensible lookup tables to support common types of transformations, such as mapping zip codes to state names for aggregation across states.
- Reshape transforms manipulate table structure and schema.
- Wrangler provides two reshaping operators: fold and unfold.
- Positional transforms include fill and lag operations.
- Fill operations generate values based on neighbouring values in a row or column and so depend on the sort order of the table.

Basic Interactions of Wrangler:

The Wrangler interface supports six basic interactions within the data table.

- Select rows
- Select columns
- Click bars in the data quality meter
- Select text within a cell, edit data values within the table
- Assign column names, data types or semantic roles.
- Choose transforms from the menu or refine suggestions by editing transform descriptions.

Application of wrangler tool :

Wrangler Interface:

The interface below shows the workshop data of staff members working in computer science department. The input data was not organized. Through split option of Wrangler tool the data was organized into tabular format.

Case Study 1: We have considered the Workshop data of Computer Science department of Dr. D. Y. Patil ACS College and applied the "Data Wrangler tool" to this data. With the help of this tool we were able to convert the data easily into tabular format. This tool provides "Wrangler transformation language with which we can easily rearrange

and visualize the data. The following screen shots show how we have applied the various Wrangler transformations: This is the first wrangler interface that we get after inserting data through Wrangler.
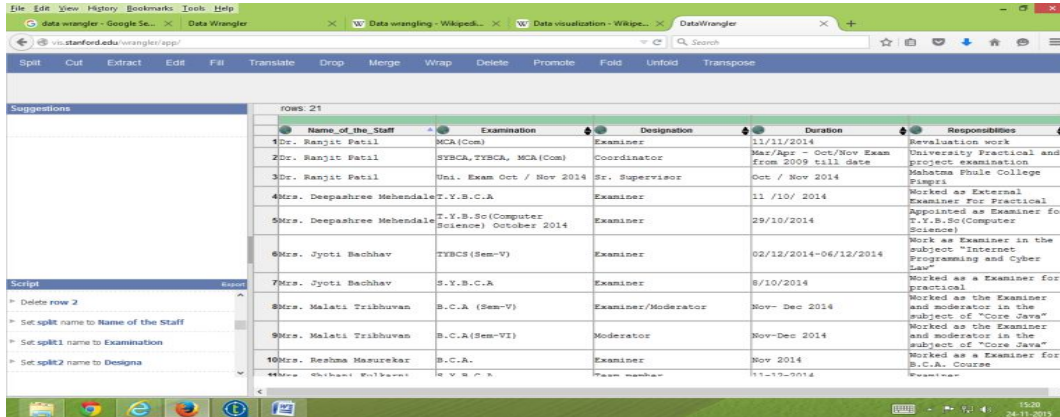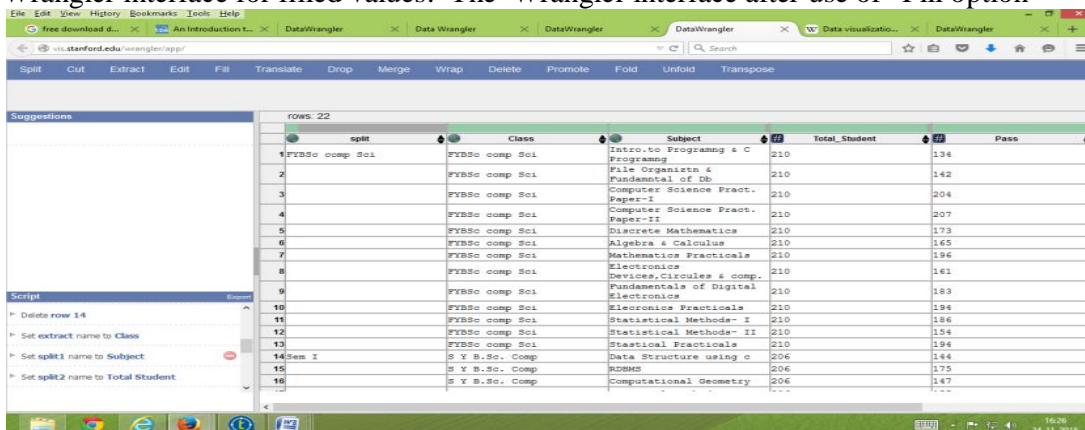

This interface shows the data for Result analysis of students. In some fields the values are missing which can be filled by using "Fill Transformation" of Wrangler.

Wrangler interface for missing values:
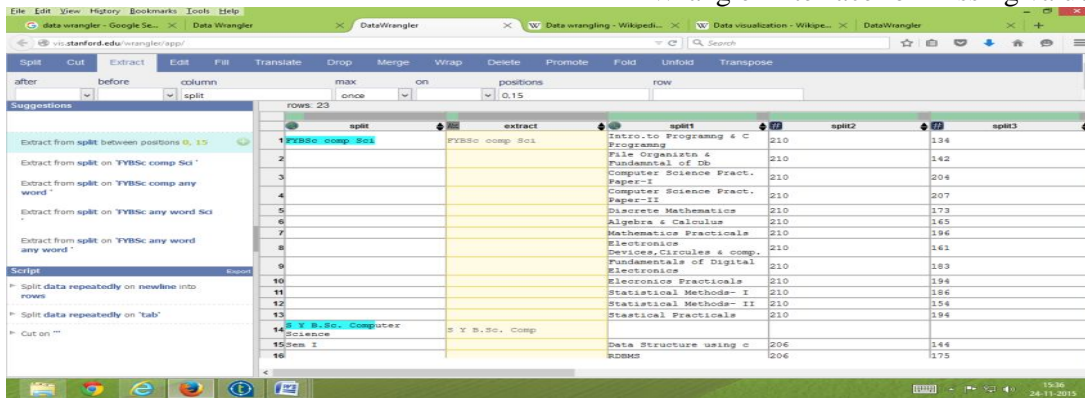
Wrangler interface for filled values: The Wrangler interface after use of "Fill option"

There are two reshape options available in wrangler i;e Fold and Unfold which we have applied.

Wrangler interface for reshape options: Fold and Unfold:

Tableau Public:

This tool can turn data into any number of visualizations, from simple to complex. One can drag and drop fields onto the work area and ask the software to suggest a visualization type, and then customize everything from labels and tool tips to size, interactive filters and legend display. Tableau Public offers a variety of ways to display interactive data. One can combine multiple connected visualizations onto a single dashboard, where one search filter can act on numerous charts, graphs and maps; underlying data tables can also be j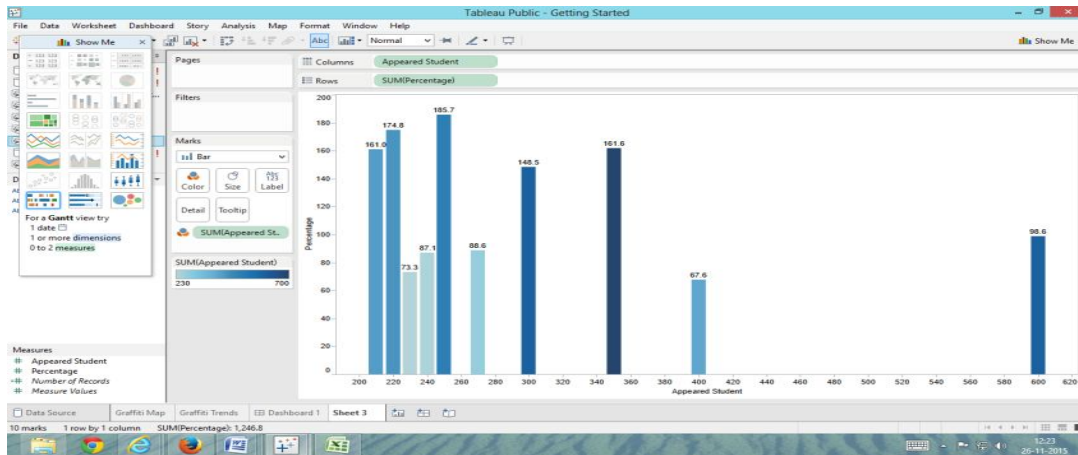oined. The tool can be better used by advanced beginner to intermediate. Tableau Public runs on Windows; native OS X version planned for 2014.

Case study 2: The second case study focuses on the use of "Tableau Public" tool .We have used this tool to generate Bar chart showing the result analysis of B.Sc. Computer science students of Dr.D.Y.Patil ACS College. Tableau tool provides various options of data visulization like Bar chart,Line Chart, Pie charts, square, circle, area, shape, polygon.

Interface for data inserted through Tableau public tool:

Bar chart generated for Result analysis of B.Sc. Computer Science students.

Conclusions:

- As mentioned above there are many free tools available in the market ,but teaching faculties at the college level are not aware of these tools. So through this paper we have tried to introduce such tools.

- All these tools are free of cost hence if we try to implement these tools in our routine work it will be cost effective and time saving.

- Wrangler is a better tool for data cleaning ,data rearrangement,data analysis and data visulization.

- Large pool of unstructured data can be easily sorted and arranged in tabular format using the data wrangler tool.

- Tableau Public is user friendly tool to convert data into graphical format.

- Though Wrangler can be used to convert data into graphical format but "Tableau Public" readily converts complex data into graphical form.

References:

- *http://vis.stanford.edu/wrangler/*
- *http://www.greenbookblog.org/2013/08/04/50-new-tools-democratizing-data-analysis-visualization/.*
- *http://www.computerworld.com/article/2507728/enterprise-applications/enterprise-applications-22-free-tools-for-data-visualization-and-analysis.html?page=4#tableau.*
- *http://www.computerworld.com/article/2502891/business-intelligence/business-intelligence-8-cool-tools-for-data-analysis-visualization-and-presentation.html?page=3.*
- *https://www.google.co.in/?gws_rd=ssl#q=tableau+public+free+tool+download.*
- *http://www.tableau.com/products/trial.*