

PROPOSAL PROYEK
12S4056 – Visualisasi Data



**VISUALIZATION FOR REVIEW BOOK BASED ON
REVIEWER**

PENGUSUL

- 12S17017 – Yolanda R.M Manurung
- 12S17027 – Stella Sitinjak
- 12S17059 – Ekis Naomi Lasma

FAKULTAS TEKNIK INFORMATIKA DAN TEKNIK ELEKTRO
INSTITUT TEKNOLOGI DEL
NOVEMBER 2020

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR GAMBAR	3
BAB I. Pendahuluan.....	5
1.1. Latar Belakang	5
1.2. Tujuan	6
1.3. Manfaat	7
1.4. Ruang Lingkup	7
BAB II. ISI.....	8
BAB III. RENCANA KERJA.....	13
3.1 Jadwal Kegiatan	13
3.2 Pembagian Tugas.....	15
BAB IV. PENUTUP	17

DAFTAR GAMBAR

Gambar 1 Arsitektur Sistem.....	8
---------------------------------	---

DAFTAR TABEL

Tabel 1 Atribut Dataset	9
Tabel 2 Jadwal Kegiatan	13
Tabel 3 Pembagian Tugas	15

BAB I. Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang pemilihan topik Proyek Visualisasi Data, tujuan yang ingin dicapai dalam pengerjaan Proyek Visualisasi Data, manfaat yang akan diperoleh, dan ruang lingkup pengerjaan.

1.1. Latar Belakang

Data merupakan sebuah kumpulan dari data objek beserta dengan atribut atau sebuah kumpulan fakta statistik yang dikumpulkan untuk dilakukan proses analisis yang tidak hanya mengacu pada satu data saja yang biasanya disimpan secara tidak terstruktur [1]. Data yang disimpan secara tidak terstruktur tersebut, apabila diolah dengan cara yang benar dan tepat, maka akan menghasilkan sebuah informasi yang dapat digunakan untuk berbagai keperluan. Salah satu contoh data yang akan digunakan dalam penelitian ini adalah data yang bersumber dari goodreads book reviews 20k yang diambil dari *book readers social network* www.goodreads.com pada tahun 2017 [2]. Data yang ditampilkan dalam dataset tersebut memiliki tipe dataset table yang terdiri dari sebuah baris (*item*), kolom (*attribute*), dan gabungan antara kolom dan baris yang dikenal dengan istilah *cell*. Namun, *user* yang ingin menggunakan data tersebut sebagai pembandingan antara buku yang satu dengan yang lain akan mengalami kesulitan dalam memutuskan buku mana yang akan dipilih untuk dibeli atau digunakan berdasarkan data yang disediakan dalam bentuk tabel tersebut.

Pada umumnya data disajikan dalam bentuk tabel dalam jumlah yang banyak dan menyebabkan *user* tidak segera mengetahui poin penting ataupun tujuan data tersebut ada. Sehingga dalam hal ini dibutuhkan sebuah visualisasi data dalam bentuk gambar ataupun bentuk lainnya untuk memudahkan *user* dalam mencerna informasi yang diberikan. Sama halnya dengan data yang akan digunakan pada penelitian ini yaitu data goodreads book reviews, jumlah data yang besar tersebut membuat *user* kewalahan dalam mengambil informasi penting didalamnya dan cenderung akan memberikan kesan malas untuk mengetahui informasi apa yang ada didalamnya. Oleh karena itu, maka sangat penting untuk merepresentasikan sebuah data ke dalam idiom atau visualisasi data [3]. Visualisasi data merupakan sebuah representasi dari informasi serta data yang ditampilkan dengan elemen atau idiom visualisasi seperti chart, grafik, map yang berguna untuk menunjukkan bagaimana gambaran umum dari data yang diperoleh untuk melakukan identifikasi terhadap data tersebut

[4]. Visualisasi data juga dapat didefinisikan sebagai rekayasa dalam pembuatan gambar untuk memberikan informasi yang secara umum bentuk dari visualisasi itu sendiri bersifat abstrak dalam bentuk skematis yang memiliki tujuan untuk mengkomunikasikan informasi secara jelas dan efisien kepada pengguna melalui grafik informasi yang telah dipilih.

Visualisasi data yang akan digunakan dalam melakukan penelitian ini diharapkan mampu untuk menjawab permasalahan *user* serta sangat memungkinkan untuk membantu *user* dalam membuat sebuah keputusan. Selain itu, berdasarkan sebuah penelitian ketika seseorang sedang melihat informasi mereka akan mengingat 80% dari apa yang dilihat dan hanya akan mengingat 20% dari apa yang dibaca. Namun tentu saja untuk membuat sebuah informasi yang mudah dipahami hanya dengan melihatnya saja bukanlah suatu hal yang mudah dilakukan. Untuk membuat suatu visualisasi yang tepat harus memperhatikan tata letak dan keseimbangan bentuk serta fungsi yang tidak akan membuat *user* berpikir keras untuk memahami informasi yang ditampilkan tersebut. Grafik yang sederhana bisa menjadi sebuah visualisasi yang sangat membosankan dan tidak memberikan informasi atau menyampaikan pesan yang tepat [5]. Sehingga dalam peneliti ini penulis akan melakukan analisis terhadap data terlebih dahulu sebelum memutuskan idiom visualisasi apa yang akan digunakan untuk merepresentasikan data tersebut. Hasil dari penentuan idiom visualisasi yang akan digunakan nantinya akan ditempatkan pada sebuah dashboard guna untuk membantu *user* untuk lebih memahami informasi yang diberikan serta membantu membuat sebuah keputusan yang tepat dan cepat berdasarkan data apa yang ditampilkan.

1.2. Tujuan

Tujuan dalam penerapan visualisasi data untuk data hasil *review* buku antara lain sebagai berikut:

1. Mengetahui persentasi review positif yang diberikan oleh *customer* setelah membeli buku
2. Mengetahui buku yang paling populer dan banyak dibeli oleh *customer*.
3. Mengetahui popularitas *author* berdasarkan rating yang diberikan *customer* terhadap buku yang dibeli.

1.3. Manfaat

Manfaat dalam penerapan visualisasi data untuk data hasil *review* buku antara lain sebagai berikut:

1. Menemukan pola, tren serta wawasan dari data hasil review sebuah buku yang akan disajikan dalam bentuk visualisasi data atau idiom
2. Membantu *user* dalam mengambil sebuah keputusan ketika akan membeli sebuah buku berdasarkan hasil review *customer* sebelumnya
3. Sistem mampu menjawab *business question* seperti buku apa yang terjual paling banyak dalam periode waktu tertentu, buku apa yang paling banyak mendapat review positif, serta buku apa yang paling diminati oleh *customer* pada waktu tertentu.

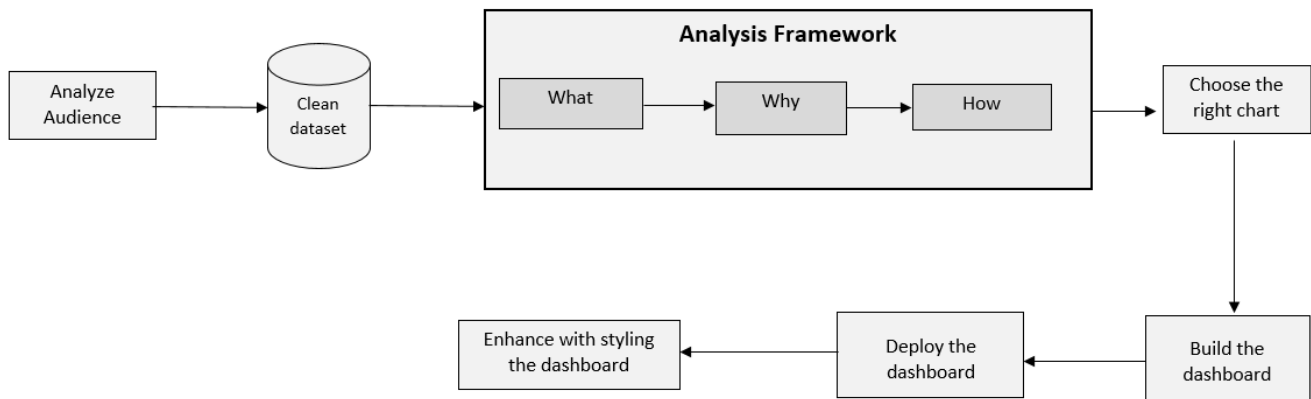
1.4. Ruang Lingkup

Ruang lingkup dalam penerapan visualisasi data untuk data hasil *review* buku antara lain sebagai berikut:

1. Sumber data berasal dari Kaggle *book readers social network* www.goodreads.com pada tahun 2017
2. Idiom akan digambarkan dengan menggunakan 3 idiom yang berbeda
3. Ketiga idiom akan divisualisasikan dalam *dashboard* menggunakan Tableau Public

BAB II. ISI

Arsitektur sistem yang akan digunakan dalam pengerjaan proyek Pemrosesan Bahasa Alami yang akan diterapkan dapat dilihat pada gambar di bawah ini:



Gambar 1 Arsitektur Sistem

Gambar di atas dapat dijelaskan secara rinci sebagai berikut:

1. *Analyze Audience*: tahap ini merupakan tahap awal yang dilakukan untuk menganalisis masalah apa yang sedang terjadi yang dapat diselesaikan dengan menggunakan sebuah visualisasi. Target *audience* merupakan faktor penting dalam membangun sebuah proyek yang akan divisualisasikan, sehingga dalam penelitian ini, penulis memutuskan untuk memilih *user* atau bahkan penulis dari sebuah buku sebagai target. Dimana, permasalahan yang terjadi akan diselesaikan dengan menggunakan sebuah visualisasi sederhana namun memberikan informasi utama.
2. *Clean dataset*: tahap ini merupakan tahapan yang akan digunakan untuk membersihkan dataset yang telah diperoleh, yakni dalam penelitian ini dataset yang akan digunakan adalah dataset goodreads book reviews 20k yang diambil dari *book readers social network* www.goodreads.com pada tahun 2017 [2]. Dataset tersebut harus dibersihkan terlebih dahulu dari data yang duplikat, sehingga tidak membuat visualisasi yang dihasilkan memberikan hasil yang tidak diharapkan dan yang tidak sesuai dengan data aslinya serta setiap baris data yang digunakan haruslah unik.

Kumpulan data tersebut berjumlah 21,6k data yang memiliki 15 atribut, yaitu:

Tabel 1 Atribut Dataset

No.	Nama Atribut	Tipe Data
1.	book_title	string
2.	book_series	string
3.	book_series_url	string
4.	book_image	string
5.	book_image_url	string
6.	book_rating	decimal
7.	book_author	string
8.	author_url	string
9.	genre	string
10.	reviewer_name	string
11.	reviewer_url	string
12.	reviewer_image	string
13.	reviewer_image_url	string
14.	review	string
15.	ID	decimal

3. *Analysis Framework:*

Analysis framework yang digunakan dalam pembangunan dashboard dalam proyek ini terbagi menjadi 3 bagian [6], antara lain:

- a. *What*: merupakan sebuah *framework* yang menandakan atribut apa yang akan digunakan dalam membuat sebuah idiom atau visualisasi data. Dalam proyek visualisasi ini, idiom yang akan dibuat yaitu terdiri dari 3 bagian dan atribut yang akan digunakan antara lain,
 1. Persentasi review positif buku: atribut yang diperlukan: genre dan book_rating

2. Presentasi buku populer, atribut yang diperlukan: *book_title*, *reviewer_name*, dan *book_rating*.
 3. Popularitas author, atribut yang diperlukan: *book_author* dan *book_rating*.
- b. *Why*: merupakan *framework* kedua dalam merepresentasikan *analysis framework* yang mendefinisikan mengapa visualisasi yang akan digambarkan pada poin satu harus dilakukan. Dalam proyek visualisasi ini, idiom yang akan dibuat yaitu terdiri dari 3 bagian sehingga alasan dalam pembuatan ketiga visualisasi tersebut adalah,
1. Visualisasi persentasi review positif yang diberikan oleh *customer* sebelumnya sangat penting, dikarenakan hasil dari review tersebut dapat digunakan sebagai perbandingan antar buku yang saling berkorelasi sehingga *user* yang akan membeli buku tidak kebingungan dalam menentukan keputusan.
 2. Visualisasi buku populer dibuat agar *user* mengetahui daftar buku-buku populer dan dapat menjadi rekomendasi untuk buku selanjutnya. Buku populer dianggap menjadi buku yang paling bagus dan banyak disukai *customer*.
 3. Visualisasi popularitas author berdasarkan rating dibutuhkan untuk dapat menjadi faktor pembanding bagi calon *customer*. Author dengan popularitas buku yang ditulisnya memiliki rating yang baik tentu saja menjadi daya tarik bagi *customer* yang akan membeli buku.
- c. *How*: merupakan *framework* ketiga dalam *analysis framework* dalam visualisasi data yang merepresentasikan bagaimana membuat sebuah idiom visualisasi data berdasarkan analisis yang sudah dilakukan pada *framework* what dan why yang sudah didefinisikan sebelumnya. Dalam proyek visualisasi ini, idiom yang akan dibuat yaitu terdiri dari 3 bagian sehingga cara untuk menggambarkan idiom tersebut antara lain:
1. Visualisasi persentasi positif yang akan dibuat yaitu dengan menggunakan pie chart, dikarenakan rating yang diberikan dalam angka decimal, sehingga akan memudahkan *user* dalam melihat persentase perbandingan dari grafik pie chart yang dibuat dalam sebuah visualisasi
 2. Visualisasi presentasi buku populer akan dibuat dengan menggunakan *bar chart* agar *user* dapat dengan mudah melihat buku yang paling disukai oleh *user* dengan hanya melihat urutan judul buku pada *chart*. *User* juga tidak perlu

melihat jumlah review atau rating dari sebuah buku, karena *chart* dapat diurutkan.

3. Visualisasi popularitas author akan dibuat dengan menggunakan idiom visualisasi bubble chart dengan menggunakan satu atribut bertipe data categorical dan sebuah atribut bertipe data quantitative.
4. *Choose the right chart*: Tahap ini merupakan tahapan untuk memilih bagaimana *chart* yang akan digunakan dalam memvisualisasikan data sesuai dengan tujuan awal dari pembangunan proyek yang telah dirancang. Data yang telah di visualisasikan diharapkan dapat dianalisis sehingga *user* atau pengguna yang melihat grafik yang ditampilkan mendapat *key point* dari grafik tersebut. Dalam proyek visualisasi ini, idiom yang akan divisualisasikan terdiri dari 3 bentuk antara lain:
 1. Idiom visualisasi yang akan digambarkan sesuai dengan tujuan awal visualisasi yaitu untuk melihat persentasi review positif dari buku yang diperoleh dari *customer* sebelumnya adalah pie chart biasa, sehingga *chart* yang dihasilkan simple dan tidak membuat bingung *user* yang melihat, bahkan *user* yang tidak memiliki pengetahuan pun bisa memahami grafik persentasi hasil review yang sudah disediakan.
 2. Bar chart yang digunakan adalah *ordered bar chart* agar *user* dapat dengan mudah melihat buku yang paling banyak dibeli oleh *user* lainnya. Penggunaan bar chart dipertimbangkan karena merupakan idiom yang paling umum dan dapat dipahami dengan mudah oleh semua kalangan pengguna.
 3. Idiom visualisasi yang digunakan adalah *bubble chart* untuk dapat memberikan informasi data yang dimaksud dan mudah dipahami oleh *user* serta bertujuan untuk menjadi pembanding antara popularitas author berdasarkan rating yang diberikan untuk buku mereka sehingga menjadi acuan untuk membeli buku tersebut.
5. *Build the dashboard*: tahap ini merupakan tahapan tujuan dari pembangunan sebuah *dashboard* yang memiliki struktur yang jelas dengan fungsi yang *user friendly* dan juga simple. Dimana *dashboard* yang akan dibangun merupakan gabungan dari ketiga *chart* yang sudah dibuat sebelumnya dan sudah didefinisikan pada tujuan awal dalam penelitian, dimana dashboard yang akan dibangun akan lebih baik sederhana, namun memberikan informasi yang jelas.
6. *Deploy the dashboard*: tahap ini merupakan tahapan untuk menyediakan dashboard dalam level yang berbeda dengan melakukan *share dashboard* yang telah dibuat pada tahap sebelumnya,

sehingga akan membuat *dashboard* yang lebih efektif dalam menyelesaikan permasalahan yang sudah didefinisikan dan tujuan awal dari pembangunan proyek yang sudah dijelaskan pada tahap-tahap sebelumnya.

7. *Enhance with styling the dashboard*: tahap ini merupakan tahapan untuk menyempurnakan *dashboard* yang telah dibuat sehingga pengguna yang melihat dashboard yang ditampilkan, tidak hanya memperoleh informasi. Tetapi juga dapat melakukan penelusuran terhadap dashboard dengan menyenangkan dan mudah, dengan melakukan formatting atau memberikan sentuhan style yang diperlukan, tetapi tetap simple dan menjaga kemudahan *user* dalam memperoleh informasi. Hal yang dapat dilakukan misalkan mengubah warna, fonts dan label yang diperlukan sesuai dengan hasil yang didiskusikan dengan pengguna yang telah melihat dashboard.

BAB III. RENCANA KERJA

3.1 Jadwal Kegiatan

Jadwal kegiatan untuk menyelesaikan proposal dapat dilihat pada tabel di bawah ini

Tabel 2 Jadwal Kegiatan

Kegiatan	Bulan																																			
	November (3) Hari					November (4) Hari						Desember (1) Hari						Desember (2) Hari						Desember (3) Hari						Desember (4) Hari			Desember (5) Hari			
	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	1	2	3	
Analyze audience																																				
Studi Literatur																																				
Mengumpulkan dataset																																				
Analysis Framework																																				
Choose right chart																																				
Build dashboard																																				
Deploy dashboard																																				
Enhance dashboard																																				
Finalisasi proyek																																				

Dalam tabel tersebut, November (3) Hari, merupakan bulan pengerjaan proyek Visualisasi Data, dimana angka 3 merupakan minggu ke-3 dalam bulan November. Hari merepresentasikan angka yang terdapat di bawah bulan, dimana hari Minggu dan hari Libur natal tidak diperhitungkan ke dalam pengerjaan proyek. Jadwal tersebut dibagi menjadi 10 kegiatan, yaitu:

1. *Analyze audience*

Pada kegiatan ini, setiap anggota kelompok memiliki waktu 3 hari pada bulan November untuk menentukan permasalahan yang dapat dijawab menggunakan visualisasi dari data yang akan digunakan. Peneliti akan menganalisis jenis pembaca yang akan menggunakan visualisasi yang dihasilkan nantinya agar permasalahan dan visualisasi yang dihasilkan sesuai.

2. *Studi literatur*

Pada kegiatan ini akan dikumpulkan semua teori yang dapat digunakan sebagai landasan dalam pembuatan visualisasi sehingga hasil yang didapatkan dapat dimengerti semua orang dan hasil yang didapatkan sesuai dengan harapan. Kegiatan ini akan dilakukan bersamaan dengan jadwal analisis audiens yang dilakukan dalam 3 hari pada bulan November

3. *Mengumpulkan dataset*

Pada kegiatan ini peneliti akan mengumpulkan semua data yang dibutuhkan untuk analisis yang didapatkan dari kegiatan pertama. Dataset yang didapatkan kemudian akan divisualisasikan untuk menjawab semua permasalahan dari hasil analisis. Kegiatan ini akan dilakukan selama 2 hari pada bulan November setelah kegiatan analisis audiens dan studi literatur selesai dilakukan.

4. *Analisis framework*

Pada kegiatan ini akan dilakukan analisis mengenai *framework* yang akan digunakan untuk membuat idiom visualisasi sehingga dapat menghasilkan visualisasi yang mudah dipahami oleh pengamat. Kegiatan ini akan dilakukan selama seminggu pada bulan November.

5. *Choose right chart*

Kegiatan ini bertujuan untuk menentukan *chart* yang sesuai untuk mempresentasikan visualisasi dari data yang dikumpulkan sehingga audiens yang menjadi target visualisasi dapat dengan mudah memahami apa yang ingin disampaikan melalui visualisasi data tersebut.

6. *Build dashboard*

Setelah setiap data selesai divisualisasikan maka setiap *chart* akan disusun untuk membuat *dashboard* sehingga audiens yang menjadi target dapat dengan mudah memahami semua idiom visualisasi data dalam satu halaman tanpa harus melalui banyak tahapan.

7. *Deploy dashboard*

Dashboard yang dibangun kemudian akan diperbaiki agar lebih efektif dari sebelumnya. Kegiatan ini membutuhkan waktu 5 hari pada bulan Desember.

8. *Enhance dashboard*

Pada kegiatan ini *dashboard* akan dianalisis kembali agar visualisasi yang ditampilkan memberikan hasil yang diharapkan. Kegiatan ini memungkinkan peneliti untuk mengurangi atau menambah fungsi yang dapat digunakan pada *dashboard* seperti filter. Kegiatan ini akan dilakukan bersamaan dengan jadwal *refine dashboard*, yaitu selama 4 hari pada bulan Desember.

9. *Refine dashboard*

Pada kegiatan ini akan dilakukan penyempurnaan dari *dashboard* yang telah dibuat dan membutuhkan waktu 4 hari. Peneliti akan memeriksa secara menyeluruh agar *dashboard* yang dihasilkan terlihat lebih rapi dan lebih mudah dipahami.

10. Finalisasi proyek

Setelah semua kegiatan selesai dilakukan, maka proyek akan difinalisasi atau disahkan sehingga dapat digunakan untuk semua pembaca.

3.2 Pembagian Tugas

Visualization Review Book based on Reviewer ini akan dikerjakan oleh 3 orang dalam satu kelompok dimana setiap anggota kelompok diberikan tanggung jawab dan tugas yang sama rata dan disesuaikan dengan minat dan kemampuan masing-masing anggota sehingga proyek dapat diselesaikan sesuai dengan hasil dan waktu yang diharapkan.

Tabel 3 Pembagian Tugas

No	Aktivitas	Penanggung Jawab	Luaran	Peran
1.	Visualisasi rating author buku	Ekis	Idiom	Ketua
2.	Visualisasi presentasi review buku	Yolanda	Idiom	Anggota
3.	Visualisasi buku yang paling banyak	Stella	Idiom	Anggota

No	Aktivitas	Penanggung Jawab	Luaran	Peran
	dibeli <i>customer</i>			
4.	Penyusunan Proposal	Ekis, Yolanda, Stella	Proposal	Tim Pmsteyol
5.	Analisis dan desain visualisasi		Desain Visualisasi	
6.	Pembuatan Dashboard		Dashboard	
7.	Penyusunan Laporan Akhir		Laporan Akhir	

BAB IV. PENUTUP

Permasalahan *user* dalam melihat sebuah data yang tidak divisualisasikan sehingga mengalami kesulitan dalam mengambil sebuah keputusan untuk memutuskan buku mana yang akan dibeli berdasarkan hasil review yang bagus, review buku apa yang memiliki nilai positif, serta bahkan akan sangat sulit bagi seorang penulis buku untuk menemukan pola seperti apa yang menjadi tren dalam masyarakat sehingga bingung dalam menuliskan genre apa yang menarik perhatian masyarakat, dapat diselesaikan dengan membuat sebuah data ke dalam bentuk visualisasi. Hasil visualisasi data yang diberikan nantinya, diharapkan mampu untuk membuat *user* dapat lebih cepat dan lebih mudah dalam memahami bagaimana tren sebuah buku dan menyelesaikan berbagai permasalahan yang sudah disebutkan sebelumnya. Oleh karena itu, penulis membuat sebuah *dashboard* yang memiliki 3 chart yang mengandung informasi utama mengenai review buku apa yang memiliki komentar positif, buku apa yang paling banyak dibeli oleh *customer*, dan siapa *author* yang paling populer berdasarkan rating buku yang diberikan dengan menggunakan 3 idiom, yaitu pie chart, bar chart, dan bubble chart. Dengan adanya sebuah *dashboard* yang ditampilkan, maka permasalahan yang telah disebutkan sebelumnya dapat diselesaikan dengan baik tanpa memakan waktu yang lama dalam memahami *dashboard* tersebut.

Referensi

- [1] J. Han, M. Kamber and J. Pei, "Data Mining: Concept and Technique," in *Data* , University of Illinois at Urbana-Champaign , 2011.
- [2] G. Apple, "Kaggle," 2018. [Online]. Available: <https://www.kaggle.com/gapple/goodreads-book-reviews-20k>. [Accessed 21 11 2020].
- [3] J. Strecker, "Data Visualization In Review," Research Award Recipient IDRC, 2012.
- [4] D. Keim, "Information Visualization and Visual Data Mining," *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 2002.
- [5] D. M. Russell, "Simple is Good: Observations of Visualization Use Amongst the Big Data Digerati," USA, Mountain View, CA 94043, 2016.
- [6] T. Munzner, "Visualization Analysis and Design," *CRC Press*, 2015.