

**Workshop visualisasi data untuk penelitian ilmu sosial dan pendidikan**

*Workshop on data visualization for research in social sciences and education*

**Dhoriva Urwatul Wutsqa, Kismiantini, Rosita Kusumawati,  
Ezra Putranda Setiawan, Bayutama Isnaini**

*Jurusan Pendidikan Matematika, Fakultas MIPA, Universitas Negeri Yogyakarta*

*\*E-mail: dhoriva\_uw@uny.ac.id*

---

**Abstrak**

Data ilmu sosial dan pendidikan di media online dapat diperoleh salah satunya dengan *data scraping*, yang dapat berupa data terstruktur dan data tidak terstruktur. Ketidakteraturan variabel-variabel yang diperoleh juga menjadi perhatian, sehingga memerlukan *data wrangling*, dengan harapan data siap dianalisis/disajikan. Penyajian data dengan tepat tentu memerlukan teknik visualisasi data yang tepat. Oleh karena itu, *workshop* visualisasi data untuk penelitian ilmu sosial dan pendidikan dengan berbantuan perangkat lunak R ini diselenggarakan. *Workshop* ini diikuti oleh mahasiswa dari berbagai program studi baik diploma, sarjana, dan pascasarjana, serta Dosen di Universitas Halu Oleo. *Workshop* ini diawali dengan pemberian materi dan demo teknik *data scraping* dan *wrangling*. Selanjutnya, peserta diberikan materi serta demo berbagai teknik visualisasi data terstruktur dan data tidak terstruktur. Berdasarkan hasil angket, pengamatan, dan tanya jawab, tampak bahwa peserta antusias serta memahami konsep dan penggunaan perangkat lunak R untuk *data scraping* dan *wrangling*, serta visualisasi data terstruktur dan tidak terstruktur.

**Kata kunci:** *Data scraping, data wrangling, visualisasi data terstruktur, visualisasi data tidak terstruktur, perangkat lunak R, penelitian ilmu sosial, penelitian ilmu pendidikan.*

**Abstract**

*Social science and education data in online media can be obtained by scraping data, which can be in the form of structured data and unstructured data. The irregularity of the variables obtained is also a concern, thus requiring data wrangling, hoping the data will be ready to be analyzed/presented. Presenting data correctly certainly requires the correct data visualization techniques. Therefore, a data visualization workshop for social science research and education with the help of R software was held. This workshop was attended by students from various diploma, undergraduate, and postgraduate study programs and lecturers at Halu Oleo University. This workshop begins with providing materials and demonstrations of data scraping and wrangling techniques. Furthermore, participants were given materials and demos of various structured and unstructured data visualization techniques. Based on the results of questionnaires, observations, and questions and answers, the participants were enthusiastic and understood the concept and use of R software for data scraping and wrangling, and visualization of structured and unstructured data.*

**Keywords:** *Data scraping, data wrangling, structured data visualization, unstructured data visualization, R software, social science research, educational science research.*

---

**PENDAHULUAN**

Penelitian ilmu sosial dan pendidikan dilakukan dalam rangka untuk memahami, menjelaskan, dan memprediksi perilaku manusia dengan mengamati, mencerminkan, dan atau mengukur fenomena sosial. Untuk dapat menjelaskan fenomena sosial dan fenomena pendidikan tentunya perlu dimiliki suatu cara agar apa yang dijelaskan dapat dengan mudah

dipahami. Cara yang dimaksud adalah visualisasi. Visualisasi dalam penelitian ilmu sosial dan pendidikan mengacu pada teknik, proses, serta alat yang memungkinkan siapapun untuk memahami dan menyajikan hasil penelitian. Selain menyajikan hasil penelitian agar lebih mudah dipahami, visualisasi data dapat digunakan untuk melihat bentuk atau pola awal sebelum melakukan analisis data. Berbagai

metode penyajian data seperti peta, grafik, dan diagram digunakan untuk menyajikan informasi dari data yang tampak kompleks dan rumit menjadi lebih mudah dipahami dan informatif. Visualisasi data adalah penggunaan alat untuk mewakili data dalam bentuk grafik, peta, plot, animasi, atau cara grafis apa pun yang membuat konten lebih mudah dipahami. Visualisasi data memberikan bentuk data, yang memungkinkan kita untuk melihat hal-hal yang sebaliknya sulit atau tidak mungkin untuk dilihat. Menurut Ramasubramanian & Singh (2017), beberapa manfaat dari visualisasi data adalah mengidentifikasi titik-titik merah dalam data dan memulai diagnostik, melacak dan mengidentifikasi hubungan antar variabel, melihat adanya *trend* dan memahami permasalahan, meringkas *spreadsheet* dan *database* yang berukuran besar, dan mudah digunakan dan sangat informatif.

Salah satu isu dalam penelitian ilmu sosial adalah data-data yang diperoleh dari media online. *Internet of things* membuat pertumbuhan data di media *online* sangat cepat. Salah satu penyimpan data yang cukup luas adalah *world wide web*. Data yang tersedia dalam *world wide web* tentu cukup besar dan beragam. Data yang cukup besar ini membuat seorang analis membutuhkan waktu dan tenaga apabila ingin mengekstrak data tertentu secara manual. Metode *scraping* data merupakan metode yang dapat mempermudah analisis untuk memperoleh data dalam berbagai bentuk dari *world wide web*. *Web scraping* adalah penggunaan alat untuk mempermudah pengumpulan sejumlah besar data *online* secara otomatis dan cepat (Bradley & James, 2019). Pengumpulan data dari web tidak mudah karena terdapat banyak teknologi yang digunakan dalam penyusunan/pendistribusian isi dari web (misalnya HTML, XML, JSON) (Boehmke, 2016).

Salah satu *packages* pada perangkat lunak R yang dapat digunakan untuk mengekstrak data dari web adalah “*rvest*”. *Packages* “*rvest*” mempunyai beberapa fungsi, yakni `read_html()`, `html_nodes()`, `html_session()`, dan `html_table()` (Krotov & Tennyson, 2021). Data yang terekstrak dapat berupa data terstruktur maupun data tidak terstruktur, yang tentu membutuhkan proses agar

data dapat dianalisis. Setelah data diperoleh dari berbagai cara salah satunya dengan teknik *web scraping*, data tidak serta merta siap untuk dianalisis, akan tetapi diperlukan proses untuk membuat data tersebut menjadi berarti atau bermakna untuk dianalisis. Proses yang dimaksud adalah *wrangling* data. *Wrangling* data adalah proses eksplorasi dan transformasi data berulang yang memungkinkan untuk dianalisis (Kandel et al., 2011). Dengan kata lain, *wrangling* data merupakan usaha dalam mengubah data (struktur atau pun tidak terstruktur) menjadi bentuk yang lebih berguna untuk dianalisis (Boehmke, 2016). *Packages* dalam perangkat lunak R yang digunakan dalam *wrangling* data diantaranya adalah “*dplyr*” dan “*tidyr*”. Oleh karena itu bahasan ini menjadi sangat penting dalam proses visualisasi data.

Dalam ilmu sosial dan pendidikan banyak ditemukan data-data yang kuantitatif yang akan lebih informatif jika disajikan secara visual. Beberapa diagram yang sering digunakan untuk mendiskripsikan data di penelitian ilmu sosial dan pendidikan adalah *box plot*, *bar chart*, *histogram*, *line graph*, *pie chart*, *scatter plot*. Visualisasi data tersebut biasanya untuk menyajikan data persentase, proporsi, atau kuantitas tertentu dari beberapa variabel atau untuk tujuan membandingkan beberapa kelompok, atau melihat hubungan antar variabel, *trend* atau bahkan, penyebaran data. Pada analitik media sosial, *word cloud* sering digunakan sebagai strategi visualisasi dari data teks (data tidak terstruktur). *World cloud* menyajikan kata-kata yang paling sering muncul dalam teks. Kabir et al. (2018) menggunakan *world cloud* untuk menganalisis teks umum dan tweet dalam mengetahui sentimen dan aspek hukum pada *text mining*. Beberapa *packages* di perangkat lunak R yang digunakan dalam membuat *world cloud* adalah “*tm*”, “*twittR*”, “*ROAuth*”, “*stringr*”, “*plyr*”, “*dplyr*”, dan “*ggplot2*”. Pada penelitian ilmu sosial, *social network analysis* merupakan salah satu analisis yang *trend* dalam proses mengeksplorasi struktur sosial. *Network diagram* adalah visualisasi yang menampilkan hubungan dan koneksi antara objek atau konsep. Suatu *network* terdiri atas titik atau simpul yang dihubungkan oleh garis yang menunjukkan

hubungan bersama. Tsubokura *et al.* (2018) menggunakan *network diagram* untuk memeriksa bagaimana isi *tweet* tentang informasi setelah kecelakaan pembangkit listrik tenaga nuklir Fukushima Daiichi disebarluaskan. Beberapa *packages* di perangkat lunak R yang digunakan dalam membuat *network diagram* adalah “igraph”, “network”, “sna”, dan “ndtv”.

Berdasarkan penjelasan tentang *scraping and wrangling*, visualisasi data dengan *box plot*, *barchart*, *histogram*, *line graph*, *pie chart*, *scatter plot*, *word cloud* dan *network diagram* dalam kaitannya dengan penelitian ilmu sosial dan pendidikan tersebut, tiga topik ini menjadi pilihan dalam kegiatan *workshop* yang dilaksanakan dalam rangka mendukung program yang sedang digencarkan oleh Universitas Negeri Yogyakarta, yaitu dosen berkegiatan di luar kampus. Program ini diharapkan dapat mempererat kerjasama dengan institusi lain ataupun masyarakat luas. Untuk mendukung program tersebut Program Studi S1 Statistika Universitas Negeri Yogyakarta menjalin kerjasama dengan Program Studi S1 Statistika Universitas Halu Oleo dengan menyelenggarakan program “*Workshop Visualisasi Data untuk Penelitian Bidang Pendidikan dan Ilmu Sosial*”. Sebagaimana yang sudah diuraikan di atas, topik ini sangat penting bagi para mahasiswa Statistika, Statistikan, dosen, dan peneliti. Program Studi S1 Statistika Universitas Halu Oleo dipilih sebagai mitra, mengingat program studi tersebut relatif baru, yaitu berdiri pada tahun 2017, sehingga masih memerlukan untuk saling bertukar ilmu dengan universitas lain. Selain itu, kurikulum di Program Studi S1 Statistika Universitas Halu Oleo tidak memuat mata kuliah visualisasi data. Padahal bahasan tentang visualisasi data sangat penting bagi para calon Statistikan. Dengan adanya kegiatan ini, selain terjalannya kerjasama dengan Program Studi S1 Statistika Universitas Halu Oleo, juga menjadi program untuk menyebarluaskan ilmu di Program Studi Statistika ke perguruan tinggi lain, sehingga membawa manfaat tidak hanya di kalangan sendiri.

## SOLUSI/TEKNOLOGI

Teknik *scraping*, *wrangling* dan visualisasi data membutuhkan alat bantu berupa perangkat lunak R. Perangkat lunak R merupakan perangkat lunak yang dapat menjalankan berbagai fungsi dalam suatu analisis, diantaranya *scraping*, *wrangling*, dan visualisasi data (Braedly dan James, 2019). Oleh karena itu, penyampaian materi dan praktik pada *workshop* ini didasari aplikasi dengan perangkat lunak R.

### *Scraping dan Wrangling Data*

Perkembangan teknologi yang berbasis internet membuat pertumbuhan data di media online bertumbuh dengan pesat. Adanya keragaman informasi di media online (*world wide web*, *www*) menyebabkan keragaman tipe data. Teknik yang dapat digunakan untuk mendapatkan data yang tersedia di media online adalah teknik *scraping* data. *Scraping* merupakan cara otomatis untuk pengumpulan data dari situs *web* (Braedly dan James, 2019). Tidak hanya data berupa angka, berbagai tipe data yang tersedia di media online seperti teks, gambar, atau video dapat diperoleh dengan menggunakan teknik *scraping* (Marres dan Weltevrede, 2013). Salah satu *packages* yang dibutuhkan dalam *scraping* data adalah *rvest* (Wickham, 2022). Beberapa fungsi dalam *rvest* ditampilkan pada Tabel 1. Berdasarkan Tabel 1, fungsi-fungsi dalam *rvest* dalam mengekstrak berbagai macam bentuk data.

Dalam melakukan *scraping* data, data yang ada dalam laman web diekstraksi. Hal ini mengakibatkan kerumitan *scraping* data bergantung pada struktur tiap halaman web. Penentuan elemen yang akan diambil/diekstrak dalam dengan mudah dipilih menggunakan *SelectorGadget*, sehingga perangkat lunak R dapat melakukan *scraping* data sesuai dengan elemen yang dimaksud.

Tabel 1. Beberapa fungsi dalam *packages rvest*

Fungsi	Deskripsi
<code>html_attr()</code>	Mendapatkan atribut elemen
<code>html_element()</code>	Memilih element dari suatu halaman HTML
<code>html_text()</code>	Mendapatkan text elemen
<code>html_table()</code>	Mengurai tabel HTML ke bentuk data frame

Data yang berasal dari berbagai sumber membuat beragamnya bentuk dari data yang terstruktur tersebut. Hal tersebut dapat menyebabkan proses analisis selanjutnya melalui visualisasi data atau pemodelan memerlukan adanya proses penyiapan data. Data wrangling merupakan teknik/proses eksplorasi dan transformasi data yang dilakukan secara berulang-ulang (Kandel *et al.*, 2011). Tahapan utama yang dilakukan dalam proses *wrangling* data diantaranya adalah melakukan diagnosa permasalahan data, menyikapi data kotor (*dirty data*), serta transformasi data. Dari tahapan-tahapan tersebut, dapat dikatakan proses inti dalam proses *wrangling* data adalah transformasi data. Tahapan transformasi data dilakukan sedemikian sehingga data menjadi lebih berguna. *Wrangling* data dapat digunakan dengan berbantuan perangkat lunak R. *Packages* pada program yang dipergunakan untuk *wrangling* data diantaranya adalah “*dplyr*” dan “*tidyverse*”. Beberapa fungsi yang terdapat dalam dua *packages* tersebut ditunjukkan pada Tabel 2.

Tabel 2. Beberapa fungsi *wrangling* data pada *packages tidyverse* dan *dplyr*.

Fungsi	Deskripsi
<code>gather()</code>	Mengumpulkan kolom menjadi baris
<code>arrange()</code>	Mengurutkan baris berdasarkan nilai kolom
<code>distinct()</code>	Menghapus baris duplikat
<code>transmute()</code>	Membuat variabel baru dengan operasi tertentu dengan menghilangkan variabel asli.
<code>full_join()</code>	Menggabungkan data
<code>group_by()</code>	Mengelompokkan data

Adapun beberapa hal yang perlu diperhatikan pada proses *wrangling* data untuk membantu membangun fondasi *wrangling* data yang kuat (Boehmke, 2016):

- 1) *Wrangling* data membutuhkan penulisan kode. Apabila peneliti merupakan pemula dalam menulis kode dengan R atau RStudio, peneliti perlu memahami beberapa bagian “command line”.
- 2) *Wrangling* data membutuhkan kemampuan untuk bekerja dengan berbagai bentuk data

seperti angka, karakter string, variabel kategorik, variabel logik, dan tanggal.

- 3) *Wrangling* data membutuhkan pengetahuan yang kuat tentang struktur data yang berbeda untuk menyimpan kumpulan data peneliti karena kumpulan data sering berisi variabel dengan panjang dan kelas yang berbeda.
- 4) Peneliti perlu memahami bagaimana mendapatkan dari sumber yang terpercaya, karena data berasal dari berbagai sumber yang mungkin mengkhawatirkan.
- 5) Peneliti harus meminimalkan duplikasi dan menulis kode yang sederhana dan mudah dibaca agar efektif dan efisien.
- 6) *Wrangling* data biasanya membutuhkan pembentukan kembali dan transformasi data yang besar agar dapat masuk ke dalam tahap visualisasi dan pemodelan.

### Visualisasi Data

Visualisasi data merupakan teknik yang digunakan untuk penyederhanaan penyampaian informasi berdasarkan data. Penyampaian informasi berarti berkaitan dengan cara komunikasi yang efektif agar informasi yang ingin disampaikan dapat dengan mudah dimengerti. Secara umum, dua filosofi dalam menggunakan teknik visualisasi data yakni kaya informasi dan minimalis (Hehman dan Xie, 2021). Kayanya informasi dalam visualisasi data berarti bahwa dari suatu alat visualisasi dapat menampilkan data sebanyak mungkin. Akan tetapi, perlu diketahui batasan subjektif terkait seberapa banyak informasi yang dapat disampaikan dalam suatu alat visualisasi sehingga tidak membuat informasi-informasi yang seharusnya mudah untuk dipahami menjadi rumit. Oleh karena itu, filosofi lain yang perlu diperhatikan adalah minimalis. Selain itu, terdapat prinsip lainnya dari memvisualisasikan data, antara lain menentukan informasi inti yang ingin disampaikan, menggunakan perangkat lunak yang tepat, menggunakan geometri yang efektif (yang sesuai kategori yang ingin disampaikan misalnya kategori jumlah, komposisi (proporsi), distribusi, atau hubungan), memberikan warna (dapat menampilkan sesuatu informasi), menyertakan suatu ukuran ketidakpastian, membuat panel (untuk dapat

membandingkan data), membedakan data dan model (data mentah biasanya lebih mudah untuk divisualisasikan), membuat visualisasi yang sederhana tetapi penjelasan/keterangan harus rinci, mempertimbangkan infografis sebagai alat visualisasi (walau tidak semua gambar dapat diubah ke bentuk infografis, tetapi infografis dikenal memberikan pemahaman tentang teks, gambar, serta elemen lainnya), meminta pendapat pihak lain terkait visualisasi yang dibuat (Midway, 2020). Visualisasi data dapat dilakukan untuk dua jenis data, yakni data terstruktur dan data tidak terstruktur.

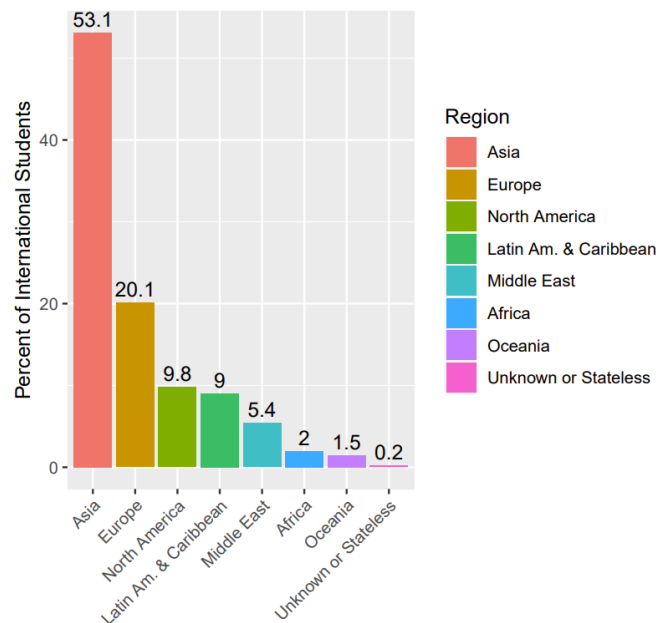
Berbagai macam teknik visualisasi data terstruktur yang dapat digunakan sesuai tujuannya masing-masing, diantaranya *box plot*, *bar chart*, *histogram*, *line graph*, *pie chart*, serta *scatter plot*. Berbagai teknik visualisasi tersebut tersedia dalam perangkat lunak R. Visualisasi sederhana dalam perangkat lunak R dapat menggunakan sejumlah fungsi yang beberapa diantaranya ditunjukkan pada Tabel 3.

Tabel 3. Beberapa fungsi dari berbagai teknik visualisasi data terstruktur dalam perangkat lunak R

Jenis diagram	Perintah R
Diagram pencar ( <i>scatter plot</i> )	plot()
Matriks diagram pencar ( <i>scatter plot matrix</i> )	pairs()
Diagram kotak garis ( <i>box plot</i> )	boxplot()
Diagram baris ( <i>strip chart</i> )	stripchart()
Histogram	hist()
Diagram densitas ( <i>density plot</i> )	density()
Diagram batang ( <i>bar chart</i> )	barplot()
Diagram garis ( <i>line plot</i> )	line() atau plot()
Diagram lingkaran ( <i>pie chart</i> )	pie()
Diagram titik	dotchart()

Selain fungsi-fungsi pada Tabel 3, terdapat *packages* lainnya dalam perangkat lunak R yang dapat digunakan untuk membuat visualisasi lebih menarik, salah satunya adalah

ggplot2 (Wickham *et al.*, 2022). Konsep dari *packages* ggplot2 adalah membagi plot dalam tiga bagian fundamental berbeda, yakni data, *aesthetic*, dan geometri. Beberapa jenis visualisasi data yang dapat dibuat menggunakan pustaka ini adalah *box plot*, *violin plot*, *dot plot*, *stripchart*, *density plot*, *histogram plot*, *scatter plot*, *bar plot*, *line plot*, *pie chart*, dan *Q-Q plot*. Contoh penggunaan ggplot2 dapat dilihat pada Gambar 1.



Gambar 1. Contoh penggunaan pustaka ggplot2 dalam visualisasi data terstruktur (diagram batang)

Data tidak hanya berjenis terstruktur, tetapi juga ada yang berjenis tidak terstruktur. Visualisasi data untuk data tak terstruktur diantaranya dapat menggunakan *word cloud* dan *network diagram*. *World cloud* menyajikan kata-kata yang paling sering muncul dalam teks. Kabir *et al.* (2018) menggunakan *world cloud* untuk menganalisis teks umum dan tweet dalam mengetahui sentimen dan aspek hukum pada *text mining*. Beberapa *packages* di perangkat lunak R yang digunakan dalam membuat *world cloud* adalah “tm”, “twittR”, “ROAuth”, “stringr”, “plyr”, “dplyr”, dan “ggplot2”. Sedangkan *network diagram* adalah visualisasi yang menampilkan hubungan dan koneksi antara objek atau konsep. Suatu *network* terdiri atas titik atau simpul yang dihubungkan oleh garis yang menunjukkan hubungan bersama. Tsubokura *et*

al. (2018) menggunakan *network diagram* untuk memeriksa bagaimana isi *tweet* tentang informasi setelah kecelakaan pembangkit listrik tenaga nuklir Fukushima Daiichi disebarluaskan. Beberapa *packages* di perangkat lunak R yang digunakan dalam membuat *network diagram* adalah “igraph”, “network”, “sna”, dan “ndtv”.

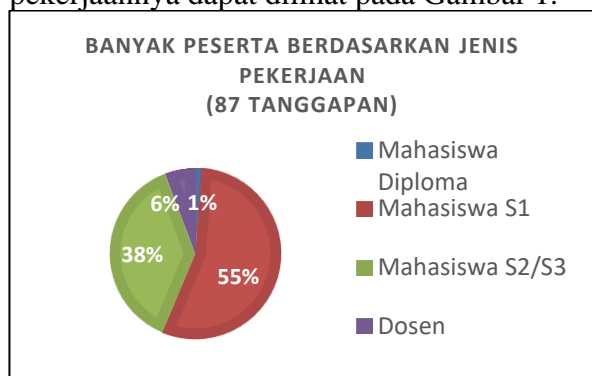
### HASIL DAN DISKUSI

Kegiatan dosen di luar kampus dalam bentuk *workshop* visualisasi data untuk penelitian da ilmu sosial dan pendidikan terbagi dalam dua tahapan, yakni tahap pra pelaksanaan dan tahap pelaksanaan.

#### Pra Pelaksanaan

Tahap pra pelaksanaan terdiri dari beberapa persiapan, yakni 1) Koordinasi waktu, bentuk kegiatan, dan sasaran kegiatan pengabdian kepada masyarakat, 2) Penyiapan materi baik teori maupun praktik, 3) Pendaftaran peserta. Pada tahap ini didapatkan kesimpulan bahwa sasaran dalam kegiatan pengabdian kepada masyarakat ini adalah mahasiswa dan dosen Universitas Halu Oleo. Pada tahap penyiapan materi dan demo program, dilakukan pembagian materi pada masing-masing narasumber untuk tiap sesi. Materi kemudian disinkronkan antar sesi sehingga dapat berkesinambungan antar materinya. Seluruh materi berupa *handout* dan file data yang diperlukan untuk mencoba program diletakkan pada folder *google drive* agar dapat diakses langsung oleh peserta pada saat pelaksanaan *workshop*.

Pada tahap pendaftaran peserta, media publikasi memanfaatkan jejaring sosial. Pada tahap ini diperoleh peminat *workshop* sebanyak 87 orang. Distribusi peserta berdasarkan jenis pekerjaannya dapat dilihat pada Gambar 1.



Gambar 2. Persentase sebaran peserta *workshop*

Berdasarkan Gambar 2, dapat dilihat bahwa mayoritas peserta yang mendaftar *workshop* merupakan mahasiswa S1 sebanyak 55% (48 orang), mahasiswa S2/S3 (Program Pascasarjana) sebanyak 38% (33 orang), dosen sebanyak 6% (5 orang), dan 1% (1 orang) merupakan mahasiswa Diploma dari Universitas Halu Oleo. Sedangkan sebaran asal program studi dari peserta ditunjukkan oleh Gambar 3.



Gambar 3. Presentasi sebaran asal program studi peserta *workshop*

Berdasarkan Gambar 3, diketahui bahwa mayoritas peserta yang mendaftar *workshop* berasal dari Program Studi Statistika yaitu sebanyak 52% (45 orang), Program Studi Pendidikan Keguruan Bahasa sebanyak 22% (19 orang), 17% (14 orang) dari Program Studi Pendidikan Matematika, 8% (7 orang) dari Program Studi Pendidikan Guru Sekolah Dasar, dan 1% (1 orang) dari Program Studi Matematika. Oleh sebab adanya program studi yang tidak familiar dengan alat pengolahan data perangkat lunak R, tim mengirimkan video terkait pengenalan dan instalasi perangkat lunak R dua hari sebelum pelaksanaan kegiatan *workshop*. Peserta dapat mempelajari video tersebut agar memiliki gambaran terkait perangkat lunak R yang akan dipergunakan pada *workshop*. Tim memanfaatkan informasi ini untuk menyusun urutan materi *workshop* dan menguji kedalaman materi yang telah dipersiapkan.

### **Pelaksanaan**

Kegiatan pengabdian kepada masyarakat yang dikemas dalam bentuk *Workshop* visualisasi data untuk penelitian ilmu sosial dan pendidikan dilaksanakan secara daring melalui zoom pada hari Kamis dan Jumat tanggal 21-22 Juli 2022. Rangkaian *workshop* dilakukan selama 14 jam yang terdiri 12 jam sesi penyampaian materi dan praktik, serta 2 jam untuk sesi tugas mandiri. Sesi penyampaian materi dan praktik dibagi ke dalam empat sesi yakni 1) sesi *scraping*, *wrangling*, dan pengantar visualisasi data, 2) sesi materi teknik visualisasi data terstruktur, 3) sesi praktik visualisasi data terstruktur, dan 4) sesi materi dan praktik teknik visualisasi data tak terstruktur. Pelaksanaan *workshop* visualisasi data untuk penelitian ilmu sosial dan pendidikan melibatkan pemateri yang merupakan anggota tim dosen berkegiatan di luar kampus serta bantuan mahasiswa dalam hal teknis persiapan.

Sesi pertama (2 jam) merupakan sesi penyampaian pengantar berbagai jenis data pada bidang ilmu sosial dan pendidikan, serta penyampaian materi dan praktik (menggunakan *scraping* data, *wrangling* data, serta pengantar visualisasi data. Pada sesi pertama juga diberikan pengantar penggunaan perangkat lunak R (instalasi). Pada sesi kedua, pemateri menyampaikan teori dari berbagai teknik visualisasi data terstruktur dan hal-hal yang perlu diperhatikan dalam melakukan visualisasi data. Sesi ketiga dilaksanakan pada hari kedua. Sesi ketiga merupakan sesi demo/praktik berbagai teknik visualisasi data terstruktur menggunakan perangkat lunak R, sedangkan untuk teori dan praktik untuk teknik visualisasi data tak terstruktur menggunakan perangkat lunak R diberikan pada sesi keempat. Peserta *workshop* diberikan tugas mandiri di akhir sesi pertama, ketiga, dan keempat. Tugas mandiri yang diberikan berupa latihan melakukan *scraping* data, *wrangling* data, serta visualisasi data terstruktur dan tak terstruktur yang berkaitan dengan data ilmu sosial dan pendidikan menggunakan perangkat lunak R. Peserta kemudian diminta mengirimkan hasilnya.

Metode yang digunakan dalam *workshop* adalah tutorial dan demo. Demo yang diberikan secara rinci membahas penerapan teori menggunakan perangkat lunak R untuk data bidang ilmu sosial dan pendidikan. Pada hari pertama, demo perangkat lunak R terkait *scraping* data dan *wrangling* data. Peserta diberikan penjelasan (demo) melakukan *scraping* data prefensi film dari *website* IMDb. Paket dalam perangkat lunak R yang dibutuhkan adalah 'xml2' dan 'rvest'. Hal-hal yang dapat didapatkan dari proses *scraping* preferensi film dari *website* IMDb tersebut diantaranya Genre, runtime, dan rating. Kemudian dilakukan demo proses *wrangling* data preferensi film tersebut. Paket yang dibutuhkan adalah 'tidyr' dan 'dplyr'. Proses *wrangling* yang dimaksud adalah membuat variabel baru diantaranya variabel kategori rating (rendah, sedang, tinggi) serta variabel anjuran berdasarkan genre masing-masing film. Pada hari kedua, juga diberikan materi tutorial maupun demo. Demo yang pertama terkait dengan teknik visualisasi data terstruktur dengan *box plot*, *bar chart*, *histogram*, *line graph*, *pie chart*, *scatter plot* untuk penelitian bidang ilmu sosial dan pendidikan. Pemateri menyampaikan bagaimana praktik teknik visualisasi data terstruktur menggunakan perangkat lunak R. Ada berbagai macam tampilan *output* teknik visualisasi data terstruktur. Misalkan untuk *scatter plot*, pemateri menyampaikan diawali dengan *basic scatter plot*. Data yang digunakan untuk praktik yaitu data world health organization (WHO) dengan format *.csv* yang berisi beberapa variabel diantaranya data nama negara, populasi, tingkat kelahiran, tingkat pendapatan negara, dan masih banyak lagi. Salah satu paket dalam perangkat lunak R yang dapat digunakan visualisasi data terstruktur adalah "ggplot2". Pada sesi kedua di hari kedua, pemateri menyampaikan materi dan praktik mengenai visualisasi data tak terstruktur dengan *word cloud* dan *network* diagram untuk penelitian bidang ilmu sosial dan pendidikan. Pada awal sesi empat, pemateri menyampaikan apa perbedaan antara data terstruktur dengan data tak terstruktur. Kemudian dilanjutkan dengan penyampaian materi dan praktik mengenai *word cloud*. Data contoh yang digunakan disimpan dalam bentuk



file *.txt*. Paket-paket dalam perangkat lunak R yang digunakan untuk membuat *word cloud* yaitu “tm”, “SnowballC”, “wordcloud”, dan “RColorBrewer”. Selanjutnya pemateri menyampaikan materi dan praktik mengenai *network graph*. Cara paling umum untuk memvisualisasikan hubungan antar entitas adalah jaringan (*network*). Jaringan atau grafik terdiri dari *node* (biasanya direpresentasikan sebagai titik) dan tepi (*edge*, biasanya direpresentasikan sebagai garis). *Packages* yang digunakan diantara yaitu “corr”, “tidyverse”, “tidygraph”. Pemateri juga menyampaikan *network graph* dengan bigram dengan *package-package* yang digunakan yaitu “igraph”, “networkD3”, “tidytext”, “tidyverse”, dan “glue”. Setelah materi dan praktik pada sesi keempat selesai disampaikan oleh pemateri, peserta *workshop* dibagi lagi menjadi 3 (tiga) kelompok dalam *breakout room*. Peserta diberi beberapa kasus dan peserta diminta untuk menyelesaikan kasus tersebut dengan mengaplikasikan ilmu yang sudah diperoleh pada sesi keempat menggunakan perangkat lunak R.

### Evaluasi

Selama kegiatan *workshop*, peserta yang hadir tampak antusias mengikuti kegiatan *workshop*. Peserta dapat mengikuti setiap perintah dalam *handout* yang diberikan untuk mengaplikasikannya pada data ilmu sosial dan pendidikan. Selain melakukan analisis berbantu perangkat lunak R, baik untuk *scraping* data, *wrangling* data, serta visualisasi data terstruktur dan tak terstruktur, peserta dapat memberikan interpretasi dari output yang diperoleh dari perangkat lunak R. Peserta terlihat cukup baik dalam mengikuti *workshop* baik pada sesi teori maupun demo perangkat lunak R. Indikasi ini terlihat dari perhatian, pertanyaan yang mereka sampaikan pada setiap sesi *workshop* khususnya pada diskusi di *breakout room*. Dengan demikian secara keseluruhan dapat disimpulkan bahwa acara *workshop* telah berjalan dengan lancar dan target acara telah tercapai.

Tabel 4. Kemampuan peserta setelah mengikuti kegiatan *workshop*

No	Kemampuan	Respon Peserta (%)			
		SS	S	TS	STS
1	Memahami cara memasang perangkat lunak R dan rstudio	43,48	56,52		
2	Memahami cara menambahkan packages yang dibutuhkan dalam perangkat lunak R/rstudio	30,43	69,57		
3	Memahami berbagai jenis data beserta teknik pengumpulannya untuk penelitian bidang pendidikan dan ilmu sosial.	34,78	65,22		
4	Memahami proses data <i>scraping</i> secara umum	30,43	69,57		
5	Melakukan data/web <i>scraping</i> menggunakan perangkat lunak R/rstudio	26,09	69,57	4,35	
6	Memahami teknik-teknik dalam proses data <i>wrangling</i>	21,74	69,57	8,70	
7	Menerapkan teknik data <i>wrangling</i> berbantu perangkat lunak R/rstudio (misalnya untuk data yang diperoleh dari hasil data <i>scraping</i> )	13,04	78,26	8,70	
8	Memahami pentingnya visualisasi data	52,17	47,83		
9	Memahami berbagai tipe visualisasi data terstruktur dan hal-hal yang perlu diperhatikan dalam melakukan visualisasi data	34,78	65,22		
10	Memahami penggunaan perangkat lunak R/rstudio untuk membuat berbagai tipe visualisasi data terstruktur	34,78	65,22		



No	Kemampuan	Respon Peserta (%)			
		SS	S	TS	STS
11	Memahami perbedaan antara data terstruktur dan data tidak terstruktur	30,43	69,57		
12	Memahami word cloud	30,43	69,57		
13	Memahami langkah-langkah dalam pembuatan <i>word cloud</i> untuk data bidang pendidikan dan ilmu sosial menggunakan perangkat lunak R/rstudio	26,09	65,22	8,70	
14	Memahami <i>network graph</i>	30,43	65,22	4,35	
15	Memahami pembuatan <i>network graph</i> untuk data bidang pendidikan dan ilmu sosial menggunakan perangkat lunak R/rstudio	17,39	78,26	4,35	
<b>Rata-rata persentase</b>		30,43	66,96	2,61	

Di akhir kegiatan workshop, panitia melakukan memberikan form evaluasi menggunakan *google form* bagi seluruh peserta. Hasil evaluasi ditunjukkan pada Tabel 4. Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 4, terlihat bahwa sebagian besar peserta merasa memiliki kemampuan dalam melakukan hal-hal yang telah diberikan dalam *workshop* selama dua hari tersebut. Tabel 4 juga memperlihatkan bahwa rata-rata presentase peserta mampu memahami konsep dan memahami aplikasi dari konsep tersebut menggunakan perangkat lunak R sebesar 30,43% menjawab Sangat Setuju, 66,96% menjawab Setuju, dan hanya 2,61% yang menjawab tidak setuju. Hal ini berarti sebagian besar peserta mampu mengikuti workshop visualisasi data untuk penelitian ilmu sosial dan pendidikan dengan baik.

## KESIMPULAN

Melalui Dosen berkegiatan di luar kampus, banyak manfaat yang dapat diberikan bagi masyarakat, dalam hal ini kelompok bidang kajian statistika, Program Studi Statiska UNY kepada mahasiswa/dosen Universitas Halu Oleo khususnya Program Studi Statistika UHO. Pengabdian kepada masyarakat yang termasuk dalam dosen berkegiatan di luar kampus yang dikemas dalam bentuk *workshop* ini, peserta mendapat pengetahuan dan keterampilan *scraping data*, *wrangling data*, visualisasi data terstruktur maupun tak terstruktur untuk penelitian di bidang ilmu sosial dan pendidikan menggunakan

perangkat lunak R. Dengan workshop ini, peserta mendapatkan gambaran betapa pentingnya data untuk dapat disajikan dengan baik agar mudah dipahami oleh orang lain. Di samping itu, kegiatan pengabdian ini juga meningkatkan kesadaran penggunaan perangkat lunak tak berbayar untuk analisis data. Tim Pengabdian kepada masyarakat (PkM) juga telah menyusun *handout* *scraping data*, *wrangling data*, serta visualisasi data terstruktur dan tak terstruktur menggunakan perangkat lunak R.

## UCAPAN TERIMA KASIH

Terima kasih kami haturkan kepada Fakultas MIPA Universitas Negeri Yogyakarta yang mendanai kegiatan PPM ini, juga kepada UPT Pusat Komputer UNY yang telah memberikan fasilitas akses Zoom untuk kegiatan ini.

## DAFTAR PUSTAKA

- Bradley, A., & James, R. J. (2019). Web scraping using R. *Advances in Methods and Practices in Psychological Science*, 2(3), 264-270.  
doi:<https://doi.org/10.1177/2515245919859535>
- Boehmke, C. B. (2016). *Data Wrangling with R*. Springer.
- Helman, E., Xie S. Y. (2021). Doing Better Data Visualization. *Advances in Methods and Practices in Psychological Science*, 4(4).  
doi:10.1177/25152459211045334

- Kabir, A. I., Karim, R., Newas, S., & Hossain, M. I. (2018). The power of social media analytics: text analytics based on sentiment analysis and word clouds on R. *Informatica Economica*, 22(1), 25-38. doi:DOI:10.12948/issn14531305/22.1.2018.03
- Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Ham, F. v., Richie, N. H., Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4), 271-288. doi:https://doi.org/10.1177/1473871611415994
- Krotov, V., & Tennyson, M. F. (2021). Web Scraping in the R Language: A Tutorial. *Journal of the Midwest Association for Information Systems*, 2021(1), 61-77. doi: 10.17705/3jmwa.000066
- Marres, N., & Weltevrede, E. (2013). Scraping the social? *Journal of Cultural Economy*, 6:3, 313-335, doi: 10.1080/17530350.2013.772070
- Midway, S. R. (2020). Principles of Effective Data Visualization. *Patterns*, 1(9). https://doi.org/10.1016/j.patter.2020.100141.
- Ramasubramanian, K., & Singh, A. (2017). Machine Learning Using R. *New Delhi: APRES*. doi:DOI:10.1007/978-1-4842-2334-5
- Tsubokura, M., Onoue, Y., Torii, H. A., Suda, S., Mori, K., Nishikawa, Y., Uno, K. (2018). Twitter use in scientific communication revealed by visualization of information spreading by influencer within half a year after the Fukushima Daiichi nuclear power plant accident. *PLoS ONE*, 13(9). doi:doi:10.1371/journal.pone.0203594
- Wickham, H. (2022). Rvest: Easily Harvest (Scrape) Web Pages. R Packages. <https://cran.r-project.org/web/packages/rvest/rvest.pdf>