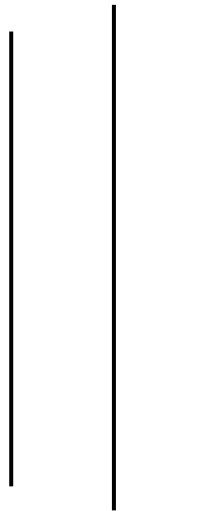


**PORTFOLIO DATA SCIENCE**  
**ANALISIS DAN VISUALISASI DATA DENGAN R STUDIO**

(Data Survey Kepuasan Terhadap Layanan Kegiatan Kemasyarakatan di Kelurahan Kota  
Administrasi Jakarta Pusat pada tahun 2018)



**DISUSUN OLEH:**  
**RINGGA ERSHA IKHWANI**

**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNOLOGI INDUSTRI**  
**UNIVERSITAS AHMAD DAHLAN**  
**YOGYAKARTA**

## A. DESKRIPSI DATA

### 1. Deskripsi Data

Pada project Data Science ini saya menggunakan Data Survey Kepuasan Terhadap Layanan Kegiatan Kemasyarakatan di Kecamatan (termasuk di Kelurahan). Kumpulan data berisi capaian Indeks Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018.

### 2. Atribut Data

Terdapat 7 atribut dari Data Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018. Sebagai berikut:

- 1) *unit\_kerja* : Nama kelurahan yang dilakukan survey
- 2) *ppsu* : Nilai indeks kepuasan masyarakat terhadap layanan PPSU (dalam persen)
- 3) *jumantik* : Nilai indeks kepuasan masyarakat terhadap layanan Jumantik (dalam persen)
- 4) *posyandu* : Nilai indeks kepuasan masyarakat terhadap layanan Posyandu (dalam persen)
- 5) *pkk* : Nilai indeks kepuasan masyarakat terhadap layanan PKK (dalam persen)
- 6) *rptra* : Nilai indeks kepuasan masyarakat terhadap layanan RPTRA (dalam persen)
- 7) *ketentraman\_ketertiban* : Nilai indeks kepuasan masyarakat terhadap layanan ketentraman dan ketertiban (dalam persen)

### 3. Jumlah Data

Jumlah data dari Data Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018 yaitu 44 observation. Dimana 44 didapatkan dari *unit\_kerja* yaitu nama kelurahan yang dilakukan survey.

### 4. Sumber Data

Data Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018 diperoleh dari [data.jakarta.go.id](http://data.jakarta.go.id) milik Organisasi Kota Administrasi Jakarta Pusat. Dimana sumber data diambil dari kelurahan kota administrasi Jakarta Pusat.

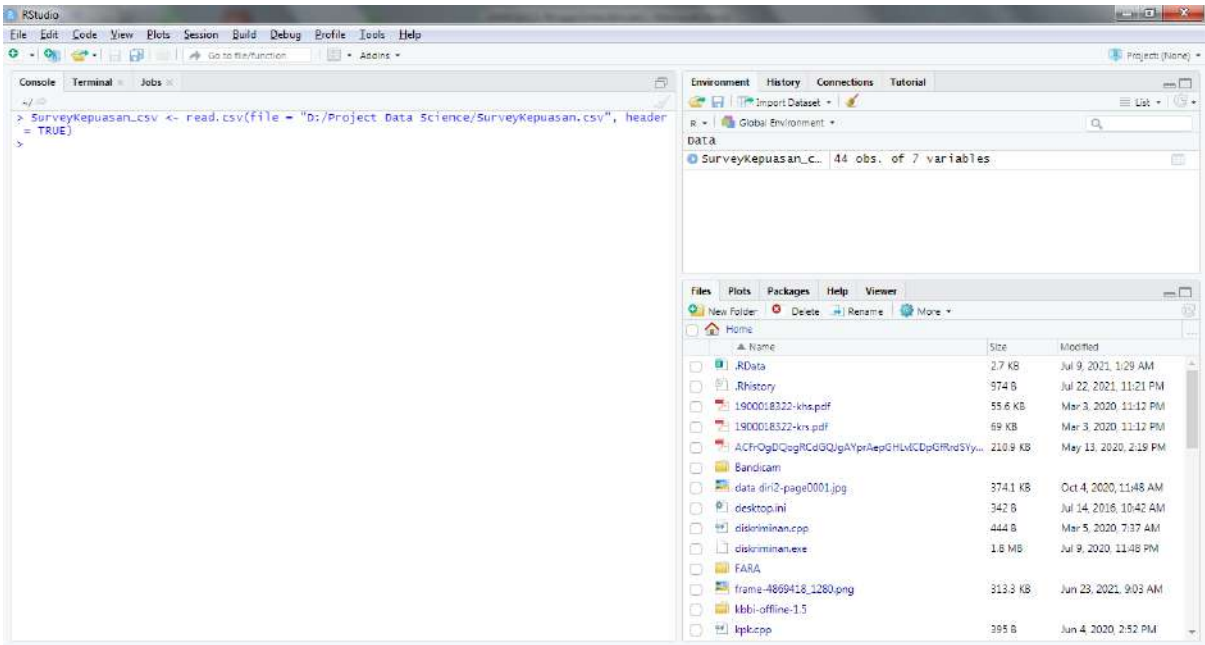
### 5. Data Collection

Pengumpulan Data (data collection) dari Data Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018 dilakukan oleh Pemerintah Kota Administrasi Jakarta Pusat.

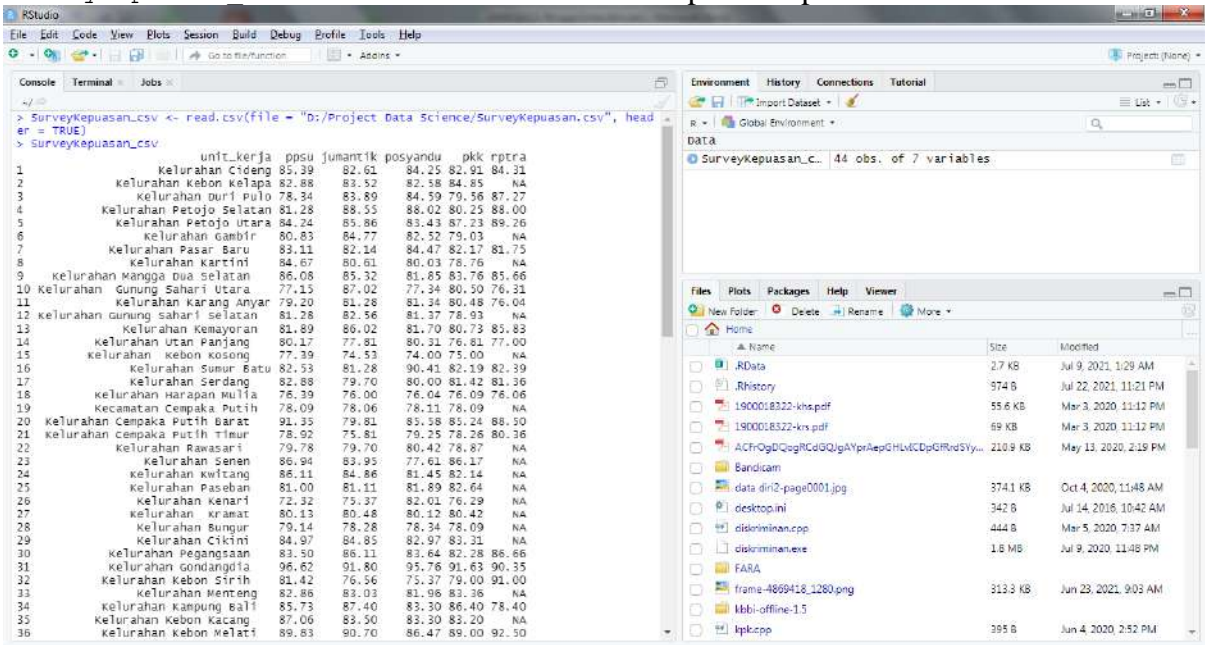
B. VISUALISASI DATA

Visualisasi Data adalah proses membuat representasi visual dari suatu data. Pada project ini saya akan membuat visualisasi dari Data Kepuasan Masyarakat di Kelurahan Kota Administrasi Jakarta Pusat pada tahun 2018. Untuk melakukan visualisasi data pada R Studio langkah yang harus dilakukan adalah sebagai berikut :

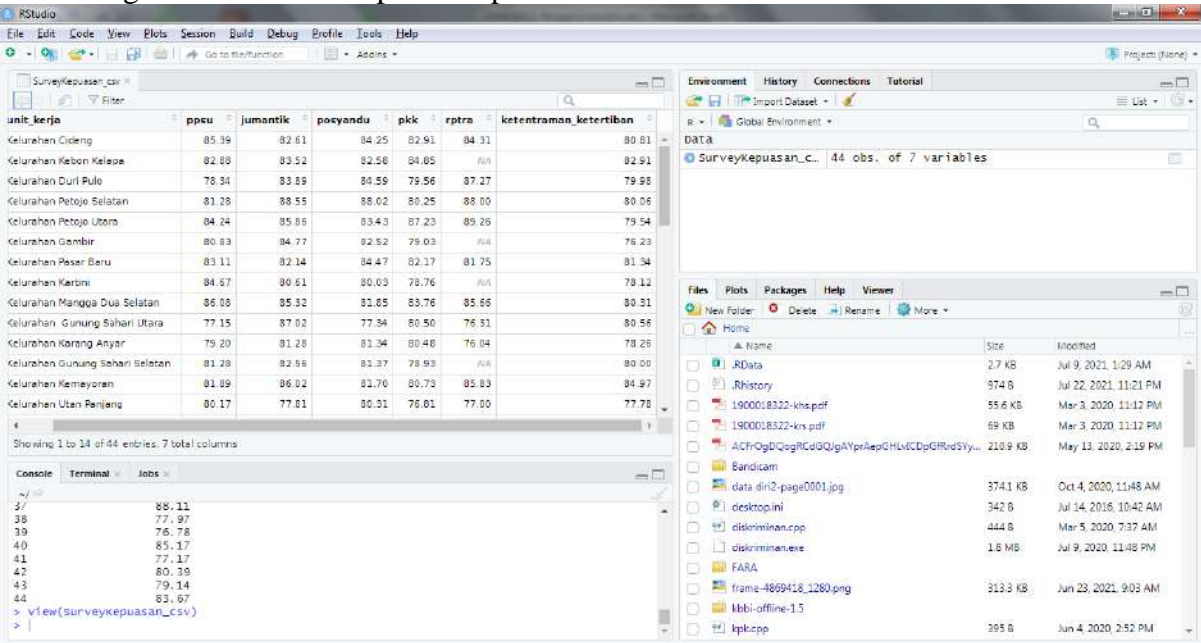
Lakukan setting working directory. disini saya mengganti nama data menjadi SurveyKepuasan\_csv. untuk melakukan pemanggilan file csv tersebut ketikkan kode : `SurveyKepuasan_csv <- read.csv(file = "D:/Project Data Science/SurveyKepuasan.csv", header = TRUE)`. Setelah itu akan muncul data SurveyKepuasan\_csv pada Global Environment dan terdapat 44 obs of 7 variables.



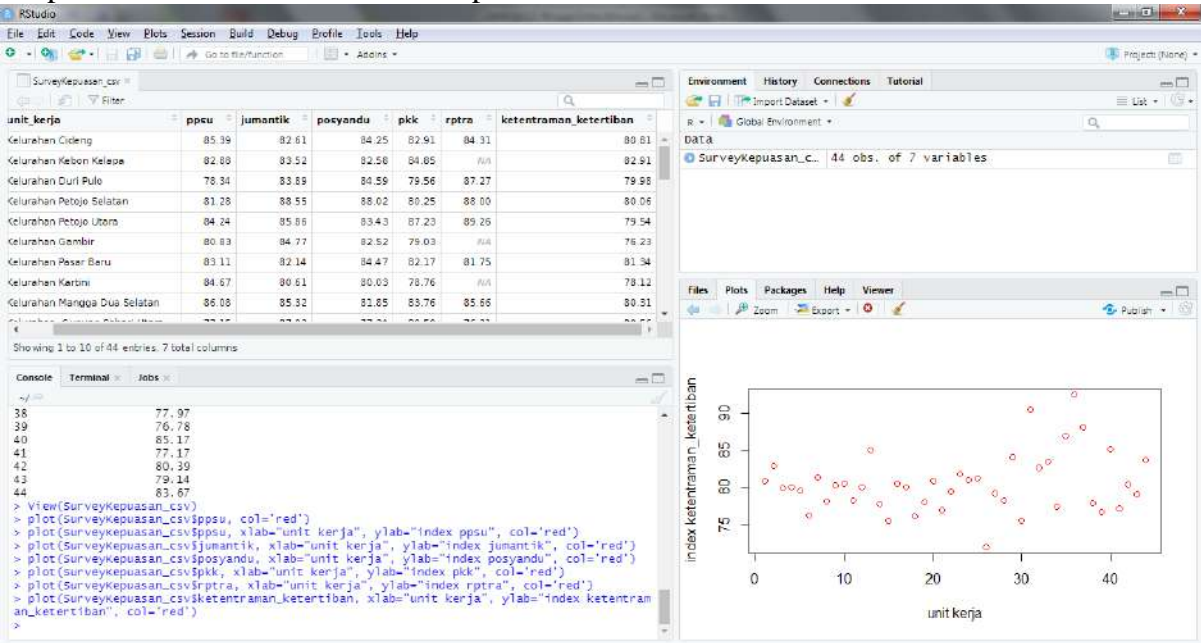
Untuk mengecek apakah file yang dipanggil sudah benar, maka ketikkan kode `SurveyKepuasan_csv`. Lalu akan muncul isi data seperti tampilan di bawah.



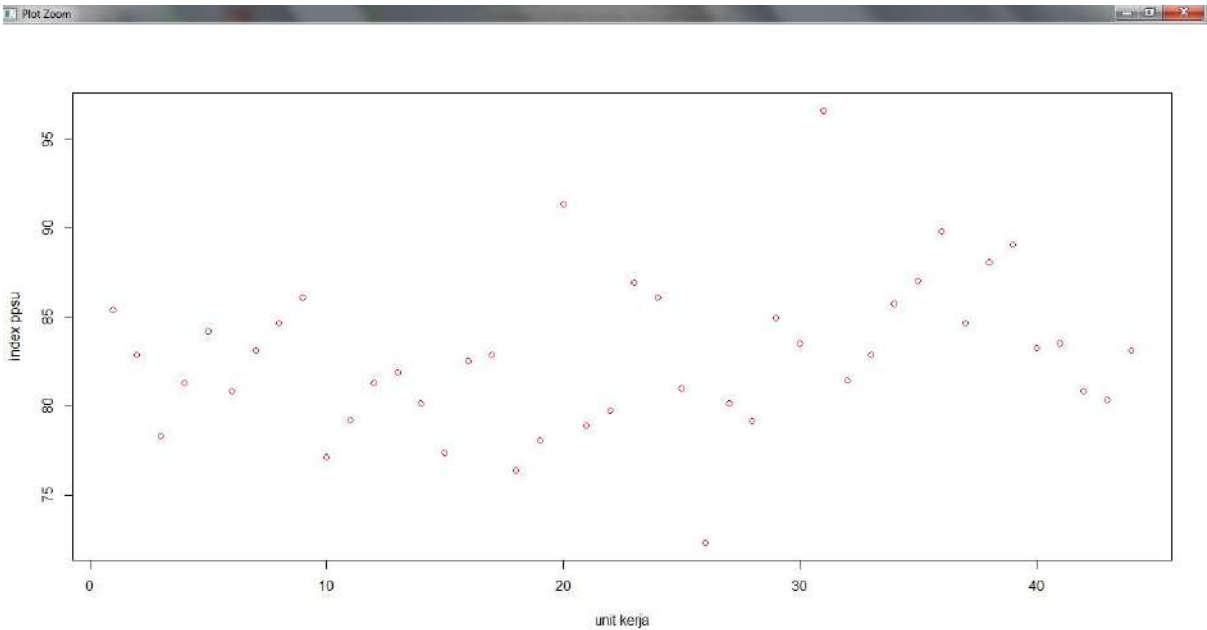
Lakukan `view(SurveyKepuasan_csv)` untuk melihat data secara jelas. Maka akan muncul data di bagian atas console seperti tampilan di bawah.



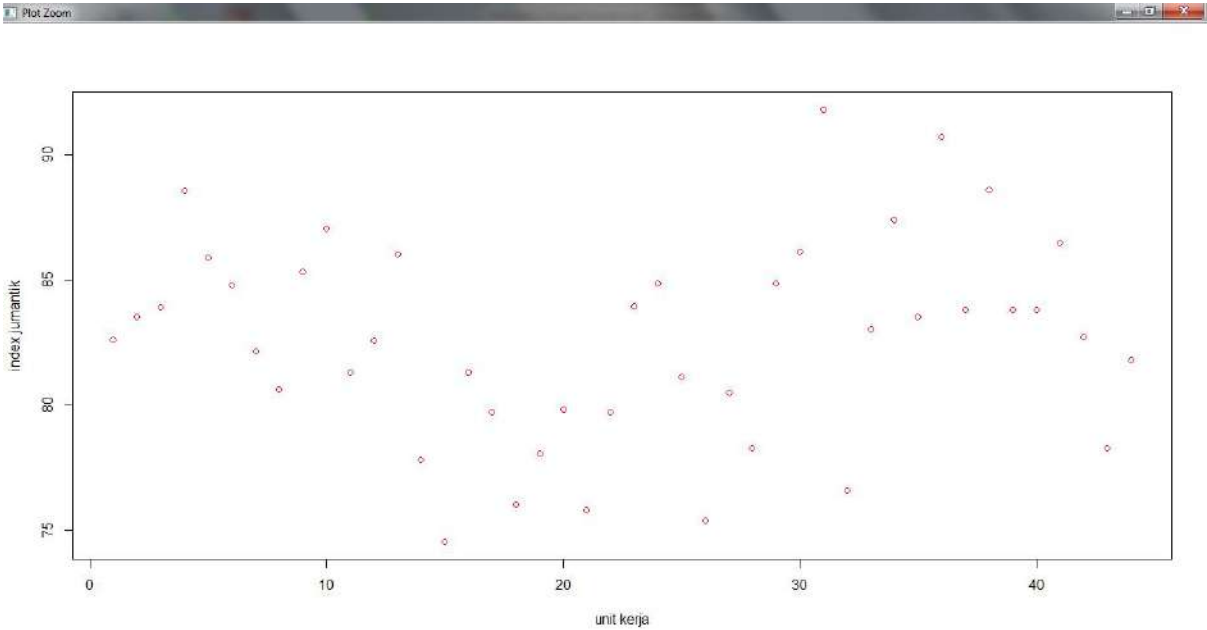
Setelah itu lakukan visualisasi data dengan menampilkan masing-masing atribut menggunakan plot. Disini saya melakukan visualisasi terhadap atribut ppsu, jumentik, posyandu, pkk, rptra, dan ketentraman\_ketertiban. dengan mengetikkan kode seperti tampilan di bawah lalu akan muncul plot.



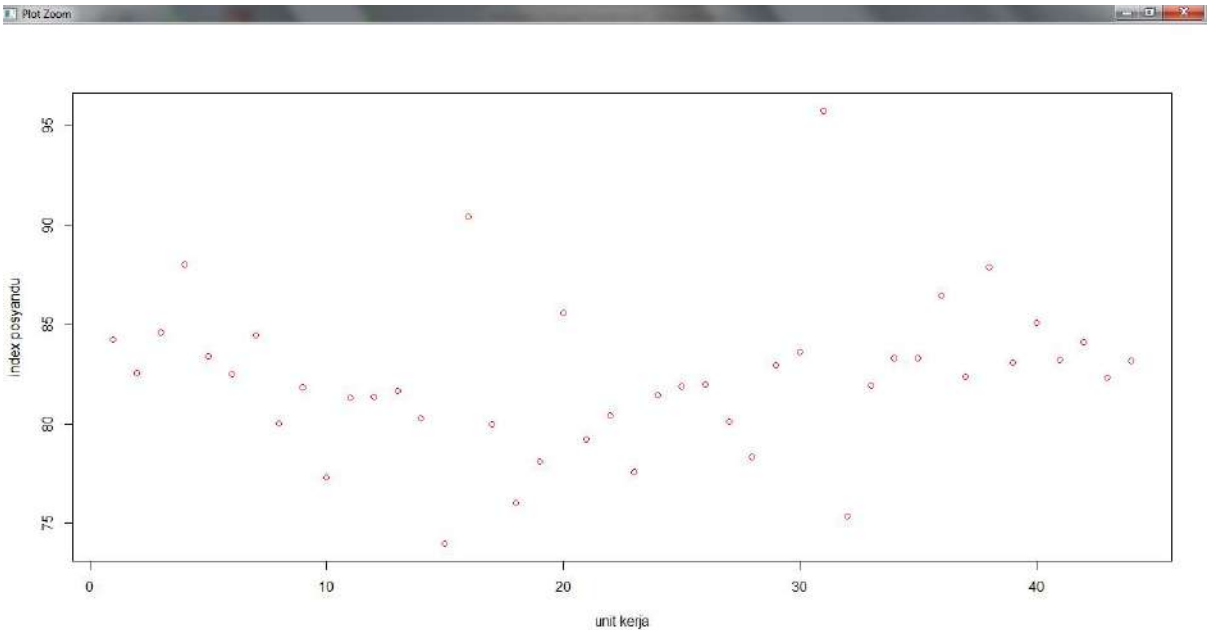
Hasil Visualisasi dari atribut ppsu



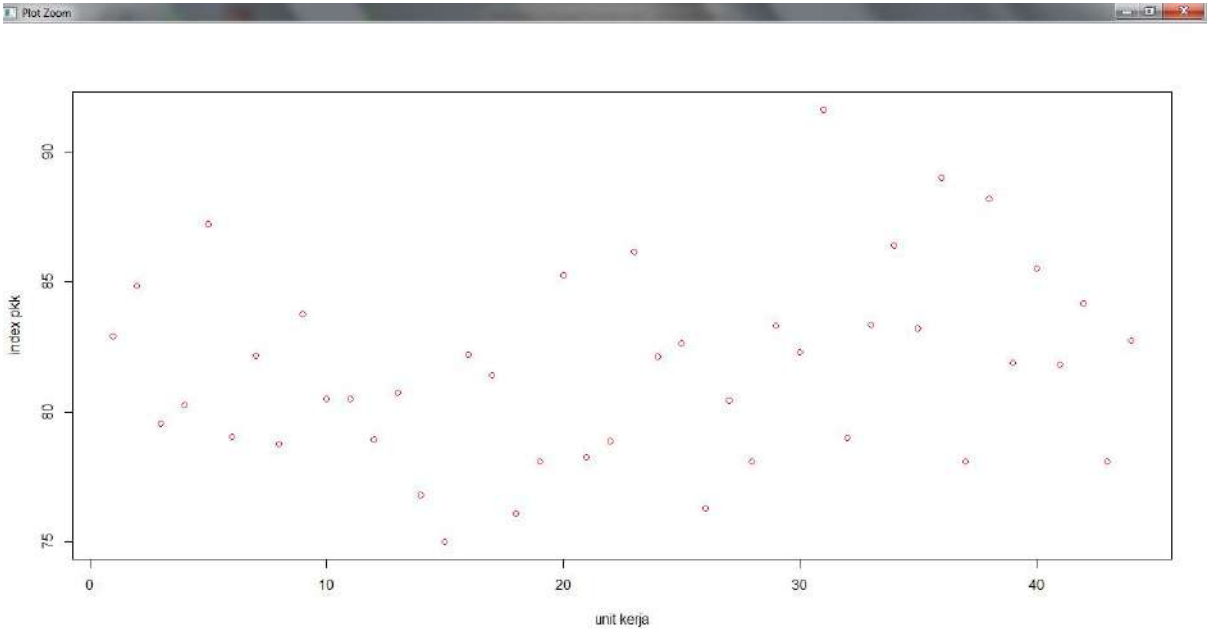
Hasil Visualisasi dari atribut jumantik



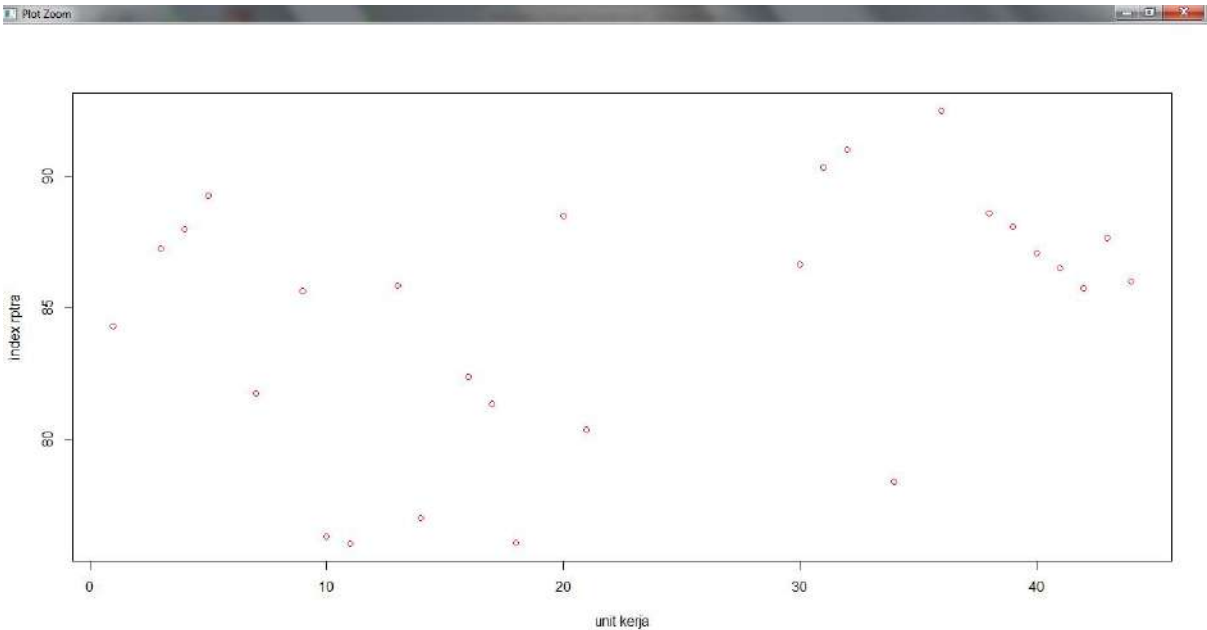
Hasil Visualisasi dari atribut posyandu



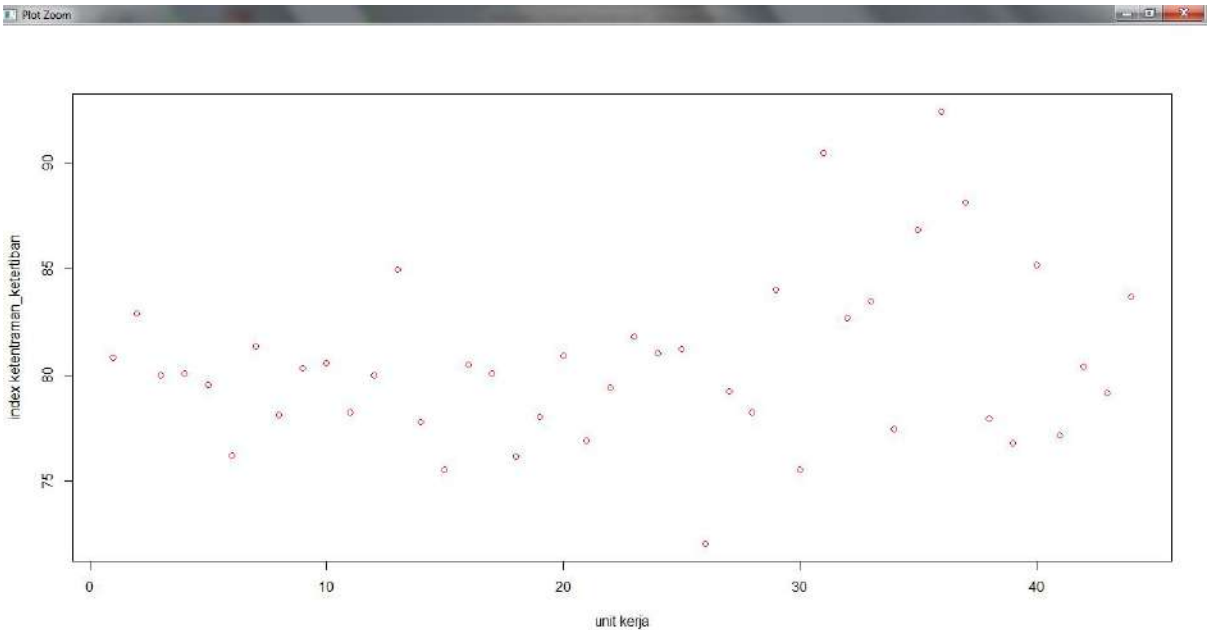
Hasil Visualisasi dari atribut pkk



Hasil Visualisasi dari atribut rptra



Hasil Visualisasi dari atribut ketentrman\_ketertiban

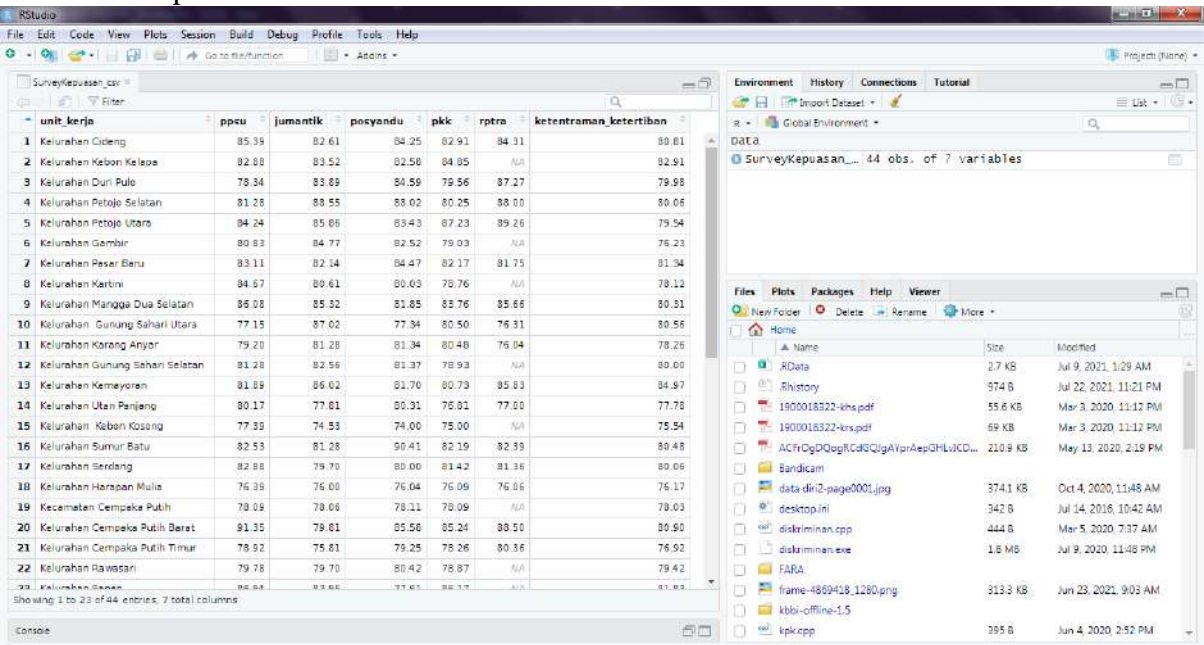




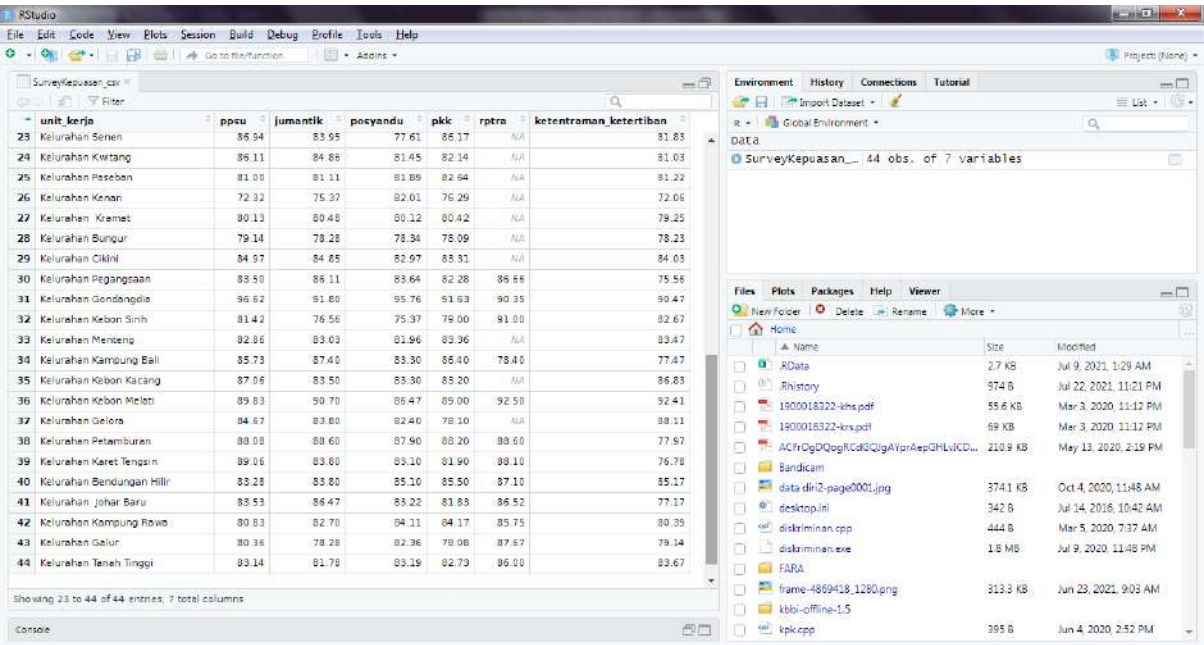
C. PREPROCESSING DATA

Preprocessing Data adalah teknik awal data mining untuk mengubah data mentah yang dikumpulkan dari berbagai sumber menjadi data yang lebih bersih dan dapat digunakan untuk pengolahan selanjutnya. Setelah melakukan visualisasi data maka selanjutnya adalah melakukan preprocessing data dari data SurveyKepuasan\_csv. Langkah awal yang dilakukan sama seperti sebelumnya yaitu ketikkan kode `SurveyKepuasan_csv <- read.csv(file = "D:/Project Data Science/SurveyKepuasan.csv", header = TRUE)`. Kemudian `View(SurveyKepuasan_csv)` Maka data akan muncul seperti tampilan di bawah.

Tampilan data SurveyKepuasan\_csv sebelum dilakukan preprocessing data. Terdapat 44 observation, dapat dilihat data masih berbentuk desimal dan terdapat beberapa missing value dari atribut `rptra`.



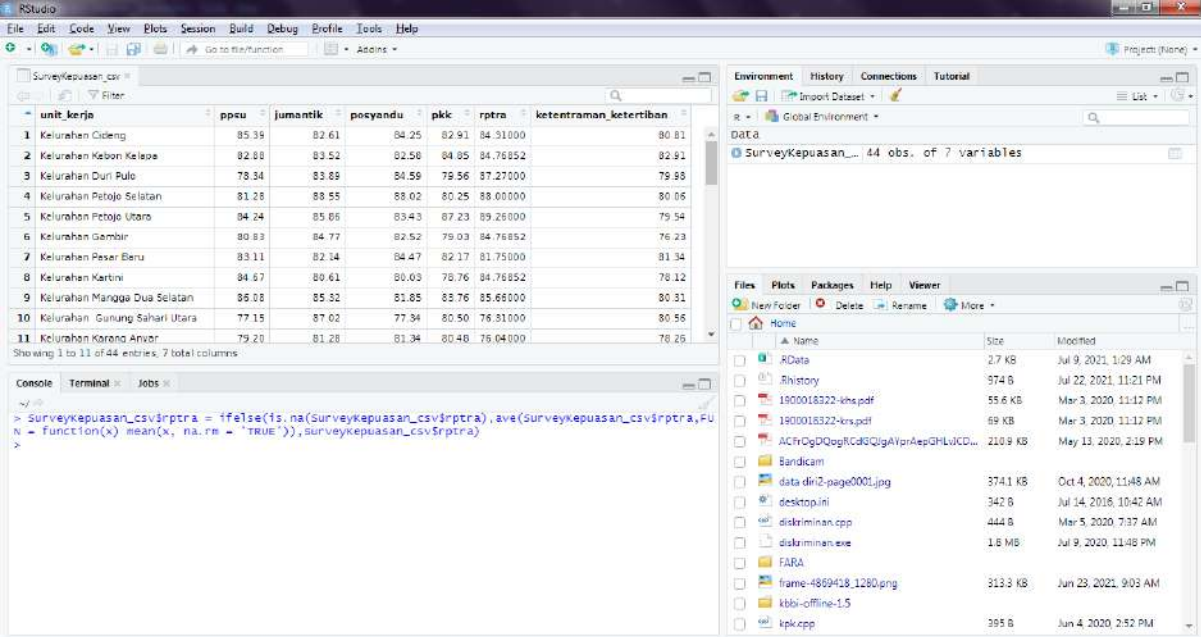
unit_kerja	ppsu	jumantik	posyandu	pkk	rptra	ketentraman_ketertiban
1 Kelurahan Cideng	85.39	82.61	84.25	82.91	84.31	80.81
2 Kelurahan Kebon Kelapa	82.88	83.52	82.58	84.05	NA	82.91
3 Kelurahan Duri Pulo	78.34	83.89	84.59	79.56	87.27	79.98
4 Kelurahan Petojo Selatan	81.28	88.55	88.02	80.25	88.00	80.08
5 Kelurahan Petojo Utara	84.24	85.86	83.43	87.23	89.26	79.54
6 Kelurahan Gembir	80.83	84.77	82.52	79.03	NA	76.23
7 Kelurahan Paser Baru	83.11	82.14	84.47	82.17	81.75	81.34
8 Kelurahan Kartini	84.67	80.61	80.03	78.76	NA	78.12
9 Kelurahan Mangga Dua Selatan	86.08	85.32	81.85	83.76	85.66	80.31
10 Kelurahan Gunung Sahari Utara	77.15	87.02	77.34	80.50	76.31	80.56
11 Kelurahan Karang Anyor	79.20	81.28	81.34	80.48	76.04	78.26
12 Kelurahan Gunung Sahari Selatan	81.28	82.56	81.37	78.93	NA	80.00
13 Kelurahan Kemayoran	81.89	86.02	81.70	80.73	85.83	84.97
14 Kelurahan Utan Panjang	80.17	77.81	80.31	76.81	77.00	77.78
15 Kelurahan Kebon Kosong	77.33	74.53	74.00	75.00	NA	75.54
16 Kelurahan Sumur Batu	82.53	81.28	90.41	82.19	82.39	80.48
17 Kelurahan Serdang	82.88	79.70	80.00	83.42	81.38	80.06
18 Kelurahan Harapan Mula	76.39	76.00	76.04	76.06	76.06	76.17
19 Kecamatan Cempaka Putih	78.09	78.06	78.11	78.09	NA	78.03
20 Kelurahan Cempaka Putih Barat	91.35	79.81	85.56	85.24	88.50	80.90
21 Kelurahan Cempaka Putih Timur	78.92	75.81	79.25	78.26	80.36	76.92
22 Kelurahan Rawasari	79.78	79.70	80.42	78.87	NA	79.42
23 Kelurahan Senen	86.88	83.86	77.61	86.17	NA	81.83



unit_kerja	ppsu	jumantik	posyandu	pkk	rptra	ketentraman_ketertiban
23 Kelurahan Senen	86.84	83.95	77.61	85.17	NA	81.83
24 Kelurahan Kertang	86.11	84.88	81.45	82.14	NA	81.03
25 Kelurahan Peseban	81.00	81.11	81.89	82.64	NA	81.22
26 Kelurahan Kenari	72.32	75.37	82.01	76.29	NA	72.06
27 Kelurahan Kramat	80.13	80.48	80.12	80.42	NA	79.25
28 Kelurahan Bungur	79.14	78.28	78.34	78.09	NA	78.23
29 Kelurahan Cikini	84.57	84.85	82.97	83.31	NA	84.03
30 Kelurahan Pegangsaan	83.50	86.11	83.64	82.28	86.86	75.56
31 Kelurahan Gondangdia	86.62	91.80	95.76	93.63	90.35	90.47
32 Kelurahan Kebon Sirih	81.42	76.56	75.37	79.00	91.00	82.67
33 Kelurahan Menteng	82.86	83.03	81.96	83.36	NA	83.47
34 Kelurahan Kampung Bali	85.73	87.40	83.30	86.40	78.40	77.47
35 Kelurahan Kebon Kacang	87.06	83.50	83.30	85.20	NA	86.83
36 Kelurahan Kebon Melati	89.83	90.70	86.47	89.00	92.50	92.41
37 Kelurahan Gelora	84.67	83.80	82.40	78.10	NA	88.11
38 Kelurahan Petamburan	88.08	88.60	87.90	88.20	88.60	77.97
39 Kelurahan Karet Tengsin	89.06	83.80	83.10	81.90	88.10	76.78
40 Kelurahan Bendungan Hilir	83.28	83.80	85.10	85.50	87.10	85.17
41 Kelurahan Johar Baru	88.53	86.47	83.22	81.83	86.52	77.17
42 Kelurahan Kampung Rawa	80.83	82.70	84.11	84.17	85.75	80.35
43 Kelurahan Galur	80.36	78.28	82.36	78.08	87.67	79.14
44 Kelurahan Tanah Tinggi	83.14	81.78	83.19	82.73	86.00	83.67



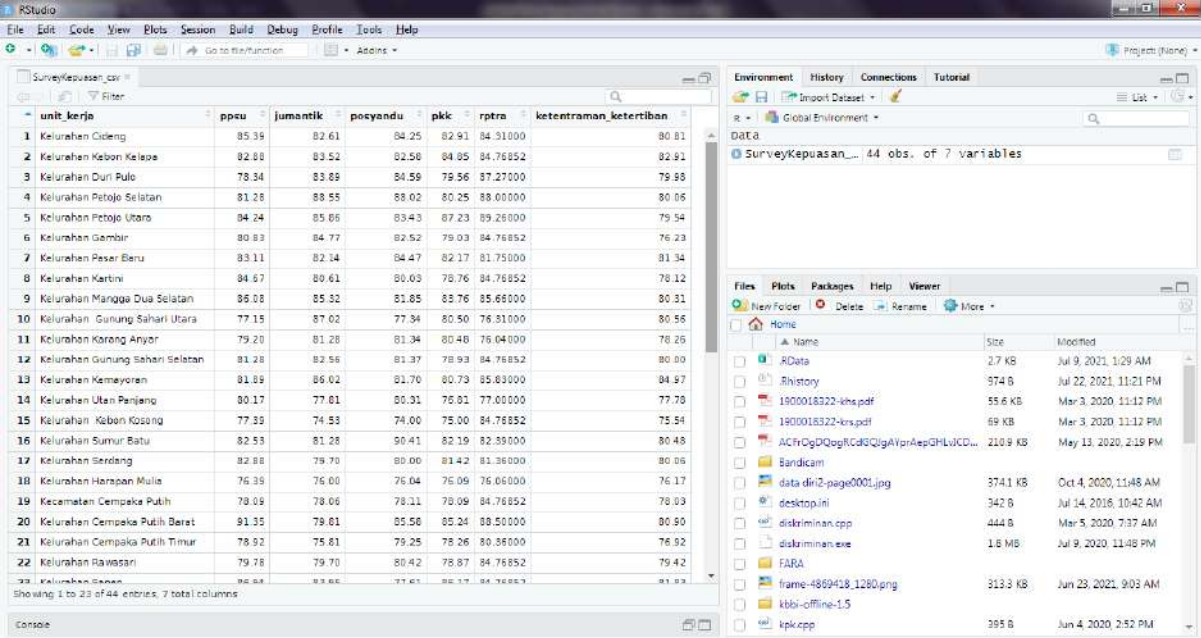
Selanjutnya yaitu lakukan handling missing value pada atribut rptra. Missing value pada atribut ini diisi dengan mean/rata-rata dari semua data dalam atribut rptra. Dengan cara ketikkan kode `SurveyKepuasan_csv$rptra = ifelse(is.na(SurveyKepuasan_csv$rptra), ave(SurveyKepuasan_csv$rptra, FUN = function(x) mean(x, na.rm = 'TRUE')), SurveyKepuasan_csv$rptra)`. Maka akan terisi nilai yang kosong seperti di bawah.



The screenshot shows the RStudio interface with the SurveyKepuasan\_csv dataset loaded. The Environment pane shows 44 observations and 7 variables. The Files pane shows the project files. The Console shows the R code used to handle missing values in the rptra column.

unit_kerja	ppsu	jumantik	posyandu	pkk	rptra	ketentraman_ketertiban
1 Kelurahan Cideng	85.39	82.61	84.25	82.91	84.51000	80.81
2 Kelurahan Kebon Kelapa	82.88	83.52	82.58	84.85	84.76852	82.91
3 Kelurahan Duri Pulo	78.34	83.89	84.59	79.56	87.27000	79.88
4 Kelurahan Petojo Selatan	81.28	88.55	88.02	80.25	88.00000	80.06
5 Kelurahan Petojo Utara	84.24	85.86	83.43	87.23	89.26000	79.54
6 Kelurahan Gambir	80.83	84.77	82.52	79.03	84.76852	76.23
7 Kelurahan Paser Baru	83.11	82.14	84.47	82.17	81.75000	81.34
8 Kelurahan Kartini	84.67	80.61	80.03	78.76	84.76852	78.12
9 Kelurahan Mangga Dua Selatan	86.08	85.32	81.85	83.76	85.66000	80.31
10 Kelurahan Gunung Sahari Utara	77.15	87.02	77.34	80.50	76.31000	80.56
11 Kelurahan Karang Anyar	79.20	81.28	81.34	80.48	79.04000	78.26

Tampilan data SurveyKepuasan\_csv setelah dilakukan handling missing value pada atribut rptra seperti di bawah.



The screenshot shows the RStudio interface with the SurveyKepuasan\_csv dataset loaded. The Environment pane shows 44 observations and 7 variables. The Files pane shows the project files. The Console shows the R code used to handle missing values in the rptra column.

unit_kerja	ppsu	jumantik	posyandu	pkk	rptra	ketentraman_ketertiban
1 Kelurahan Cideng	85.39	82.61	84.25	82.91	84.51000	80.81
2 Kelurahan Kebon Kelapa	82.88	83.52	82.58	84.85	84.76852	82.91
3 Kelurahan Duri Pulo	78.34	83.89	84.59	79.56	87.27000	79.88
4 Kelurahan Petojo Selatan	81.28	88.55	88.02	80.25	88.00000	80.06
5 Kelurahan Petojo Utara	84.24	85.86	83.43	87.23	89.26000	79.54
6 Kelurahan Gambir	80.83	84.77	82.52	79.03	84.76852	76.23
7 Kelurahan Paser Baru	83.11	82.14	84.47	82.17	81.75000	81.34
8 Kelurahan Kartini	84.67	80.61	80.03	78.76	84.76852	78.12
9 Kelurahan Mangga Dua Selatan	86.08	85.32	81.85	83.76	85.66000	80.31
10 Kelurahan Gunung Sahari Utara	77.15	87.02	77.34	80.50	76.31000	80.56
11 Kelurahan Karang Anyar	79.20	81.28	81.34	80.48	79.04000	78.26
12 Kelurahan Gunung Sahari Selatan	81.28	82.56	81.37	78.93	84.76852	80.00
13 Kelurahan Kemayoran	81.89	86.02	81.70	80.73	85.83000	84.97
14 Kelurahan Uten Panjang	80.17	77.01	80.31	76.01	77.00000	77.78
15 Kelurahan Kebon Kosong	77.33	74.53	74.00	75.00	84.76852	75.54
16 Kelurahan Sumur Batu	82.53	81.28	90.41	82.19	82.39000	80.48
17 Kelurahan Beresang	82.88	79.70	80.00	81.42	81.36000	80.06
18 Kelurahan Harapan Mulia	76.39	76.00	76.04	76.09	76.06000	76.17
19 Kecamatan Cempaka Putih	78.09	78.06	78.11	78.09	84.76852	78.03
20 Kelurahan Cempaka Putih Barat	91.35	79.81	85.56	85.24	88.50000	80.80
21 Kelurahan Cempaka Putih Timur	78.52	75.81	79.25	78.26	80.36000	76.92
22 Kelurahan Rawasari	79.78	79.70	80.42	78.87	84.76852	79.42
23 Kelurahan Cempaka Putih	82.84	83.86	77.61	80.17	84.76852	81.83

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

SurveyKepuasan\_csv

unit\_kerja ppsu jumantik posyandu pkk rptra ketentraman\_ketertiban

23	Kelurahan Senen	86.94	83.95	77.61	86.17	84.76852	81.83
24	Kelurahan Kertang	86.11	84.86	81.45	82.14	84.76852	81.03
25	Kelurahan Paseban	81.00	81.11	81.89	82.64	84.76852	81.22
26	Kelurahan Kenari	72.32	75.37	82.01	76.29	84.76852	72.06
27	Kelurahan Kramat	80.13	80.48	80.12	80.42	84.76852	79.25
28	Kelurahan Bungur	79.14	78.28	78.34	78.09	84.76852	78.23
29	Kelurahan Cikini	84.97	84.85	82.97	83.31	84.76852	84.83
30	Kelurahan Pegangsaan	83.50	86.11	83.64	82.28	86.68000	75.56
31	Kelurahan Gondangdia	86.62	91.80	95.76	93.63	80.35000	90.47
32	Kelurahan Kebon Sirih	81.42	76.56	75.37	79.00	91.00000	82.67
33	Kelurahan Menteng	82.86	83.03	81.96	83.36	84.76852	83.47
34	Kelurahan Kampung Bali	85.73	87.40	83.30	86.40	78.40000	77.47
35	Kelurahan Kebon Kacang	87.06	83.50	83.30	83.20	84.76852	96.83
36	Kelurahan Kebon Melati	89.83	90.70	86.47	89.00	92.50000	92.41
37	Kelurahan Gelora	84.67	83.80	82.40	78.10	84.76852	88.11
38	Kelurahan Petamburan	88.08	88.60	87.90	88.20	88.60000	77.97
39	Kelurahan Karet Tengsin	89.06	83.80	83.10	81.90	88.10000	76.78
40	Kelurahan Bendungan Hilir	83.28	83.80	85.10	85.50	87.10000	85.17
41	Kelurahan Johar Baru	83.53	86.47	83.22	81.83	86.52000	77.17
42	Kelurahan Kampung Rawa	80.83	82.70	84.11	84.17	85.75000	80.39
43	Kelurahan Gelur	80.36	78.28	82.36	78.06	87.67000	79.14
44	Kelurahan Tanah Tinggi	83.14	81.78	83.19	82.73	86.00000	83.67

Showing 22 to 44 of 44 entries, 7 total columns

Environment History Connections Tutorial

Global Environment

Data

SurveyKepuasan\_... 44 obs. of 7 variables

Files Plots Packages Help Viewer

Home

Name	Size	Modified
RData	2.7 KB	Jul 9, 2021, 1:29 AM
Rhistory	974 B	Jul 22, 2021, 11:21 PM
1900018322-khs.pdf	55.6 KB	Mar 3, 2020, 11:12 PM
1900018322-krs.pdf	69 KB	Mar 3, 2020, 11:12 PM
ACFrOgDQagRcd8QlgAVonAepGHLuCD...	210.9 KB	May 13, 2020, 2:19 PM
Bandicam		
data dir2-page0001.jpg	374.1 KB	Oct 4, 2020, 11:48 AM
desktop.ini	342 B	Jul 14, 2018, 10:42 AM
diskriminan.cpp	444 B	Mar 5, 2020, 7:37 AM
diskriminan.exe	1.6 MB	Jul 9, 2020, 11:48 PM
FARA		
frame-4869428.1280.png	313.3 KB	Jun 23, 2021, 9:03 AM
kbbi-offline-1.5		
kpk.cpp	395 B	Jun 4, 2020, 2:52 PM

Tahap selanjutnya yaitu melakukan pembulatan pada masing-masing atribut. dengan cara ketikkan kode berikut :

- **Pembulatan atribut ppsu :**  
`SurveyKepuasan_csv$ppsu = as.numeric(format(round(SurveyKepuasan_csv$ppsu, 0)))`
- **Pembulatan atribut jumantik :**  
`SurveyKepuasan_csv$jumantik=as.numeric (format(round(SurveyKepuasan_csv$jumantik, 0)))`
- **Pembulatan atribut posyandu :**  
`SurveyKepuasan_csv$posyandu= as.numeric(format(round(SurveyKepuasan_csv$posyandu, 0)))`
- **Pembulatan atribut pkk :**  
`SurveyKepuasan_csv$pkk = as.numeric(format(round(SurveyKepuasan_csv$pkk, 0)))`
- **Pembulatan atribut rptra :**  
`SurveyKepuasan_csv$rptra = as.numeric(format(round(SurveyKepuasan_csv$rptra, 0)))`
- **Pembulatan atribut ketentraman\_ketertiban :**  
`SurveyKepuasan_csv$ketentraman_ketertiban=as.numeric(format(round(SurveyKepuasan_csv$ketentraman_ketertiban, 0)))`

Seperti tampilan di bawah ini.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

SurveyKepuasan\_csv

unit\_kerja ppsu jumantik posyandu pkk rptra ketentraman\_ketertiban

1	Kelurahan Cideng	85	83	84	83	84	81
2	Kelurahan Kebon Kelapa	83	84	83	85	85	83
3	Kelurahan Duri Pulo	78	84	85	80	87	80
4	Kelurahan Petojo Selatan	81	89	88	80	88	80
5	Kelurahan Petojo Utara	84	86	83	87	89	80
6	Kelurahan Gambir	81	85	83	79	85	76
7	Kelurahan Pasar Baru	83	82	84	82	82	81
8	Kelurahan Kartini	85	81	80	79	85	78

Showing 1 to 8 of 44 entries, 7 total columns

Environment History Connections Tutorial

Global Environment

Data

SurveyKepuasan\_... 44 obs. of 7 variables

Files Plots Packages Help Viewer

Home

Name	Size	Modified
RData	2.7 KB	Jul 9, 2021, 1:29 AM
Rhistory	974 B	Jul 22, 2021, 11:21 PM
1900018322-khs.pdf	55.6 KB	Mar 3, 2020, 11:12 PM
1900018322-krs.pdf	69 KB	Mar 3, 2020, 11:12 PM
ACFrOgDQagRcd8QlgAVonAepGHLuCD...	210.9 KB	May 13, 2020, 2:19 PM
Bandicam		
data dir2-page0001.jpg	374.1 KB	Oct 4, 2020, 11:48 AM
desktop.ini	342 B	Jul 14, 2018, 10:42 AM
diskriminan.cpp	444 B	Mar 5, 2020, 7:37 AM
diskriminan.exe	1.6 MB	Jul 9, 2020, 11:48 PM
FARA		
frame-4869428.1280.png	313.3 KB	Jun 23, 2021, 9:03 AM
kbbi-offline-1.5		
kpk.cpp	395 B	Jun 4, 2020, 2:52 PM

Console Terminal Jobs

```
> surveykepuasan_csv$rptra = ifelse(is.na(surveykepuasan_csv$rptra), ave(surveykepuasan_csv$rptra, FUN = function(x) mean(x, na.rm = TRUE))), surveykepuasan_csv$rptra)
> SurveyKepuasan_csv$ppsu = as.numeric(format(round(SurveyKepuasan_csv$ppsu, 0)))
> SurveyKepuasan_csv$jumantik = as.numeric(format(round(SurveyKepuasan_csv$jumantik, 0)))
> SurveyKepuasan_csv$posyandu = as.numeric(format(round(SurveyKepuasan_csv$posyandu, 0)))
> SurveyKepuasan_csv$pkk = as.numeric(format(round(SurveyKepuasan_csv$pkk, 0)))
> SurveyKepuasan_csv$rptra = as.numeric(format(round(SurveyKepuasan_csv$rptra, 0)))
> SurveyKepuasan_csv$ketentraman_ketertiban = as.numeric(format(round(SurveyKepuasan_csv$ketentraman_ketertiban, 0)))
> |
```



Tampilan akhir data SurveyKepuasan\_csv setelah dilakukan preprocessing data pada masing-masing atribut. sudah tidak ada missing value dan semua atribut dalam bentuk integer(bilangan bulat) dapat dilihat seperti tampilan di bawah.

	unit_kerja	ppsu	jumatik	posyandu	pkk	rptra	ketentraman_ketertiban
1	Kelurahan Cideng	85	83	84	83	84	81
2	Kelurahan Kebon Kelapa	83	84	83	85	85	83
3	Kelurahan Duri Pulo	78	84	85	80	87	80
4	Kelurahan Petojo Selatan	81	85	88	80	88	80
5	Kelurahan Petojo Utara	84	86	83	87	85	80
6	Kelurahan Gambir	81	85	83	79	85	76
7	Kelurahan Pesir Baru	83	82	84	82	82	81
8	Kelurahan Kartini	85	81	80	79	85	78
9	Kelurahan Mangga Dua Selatan	86	85	82	84	86	80
10	Kelurahan Gunung Sahari Utara	77	87	77	80	76	81
11	Kelurahan Karang Anyar	75	81	81	80	76	78
12	Kelurahan Gunung Sahari Selatan	81	83	81	79	85	80
13	Kelurahan Kemayoran	82	86	82	81	86	85
14	Kelurahan Utan Panjang	80	78	80	77	77	78
15	Kelurahan Kebon Kosong	77	75	74	75	85	76
16	Kelurahan Sumur Batu	83	81	90	82	82	80
17	Kelurahan Serdang	83	80	80	81	81	80
18	Kelurahan Harapan Mulia	76	76	76	76	76	76
19	Kecamatan Cempaka Putih	78	78	78	78	85	78
20	Kelurahan Cempaka Putih Barat	91	80	86	85	88	81
21	Kelurahan Cempaka Putih Timur	79	76	79	78	80	77
22	Kelurahan Rawasari	80	80	80	79	85	79

	unit_kerja	ppsu	jumatik	posyandu	pkk	rptra	ketentraman_ketertiban
23	Kelurahan Senen	87	84	78	86	85	82
24	Kelurahan Kertang	86	85	81	82	85	81
25	Kelurahan Paseban	81	81	82	83	85	81
26	Kelurahan Kenari	72	75	82	76	85	72
27	Kelurahan Kramat	80	80	80	80	85	79
28	Kelurahan Bungur	75	78	76	78	85	78
29	Kelurahan Cikini	85	85	83	83	85	84
30	Kelurahan Pegangsaan	84	86	84	82	87	76
31	Kelurahan Gondangdia	97	92	96	92	90	90
32	Kelurahan Kebon Sirih	81	77	75	79	81	83
33	Kelurahan Menteng	83	83	82	83	85	83
34	Kelurahan Kampung Bali	88	87	85	86	78	77
35	Kelurahan Kebon Kacang	87	84	83	83	85	87
36	Kelurahan Kebon Melati	90	91	86	89	92	92
37	Kelurahan Gelora	85	84	82	76	85	88
38	Kelurahan Petamburan	88	89	88	88	88	78
39	Kelurahan Karet Tengsin	89	84	83	82	88	77
40	Kelurahan Bendungan Hilir	83	84	85	86	87	85
41	Kelurahan Johar Baru	84	86	83	82	87	77
42	Kelurahan Kampung Rawa	81	83	84	84	86	80
43	Kelurahan Gelur	80	78	82	76	88	79
44	Kelurahan Tanah Tinggi	83	82	83	83	88	84

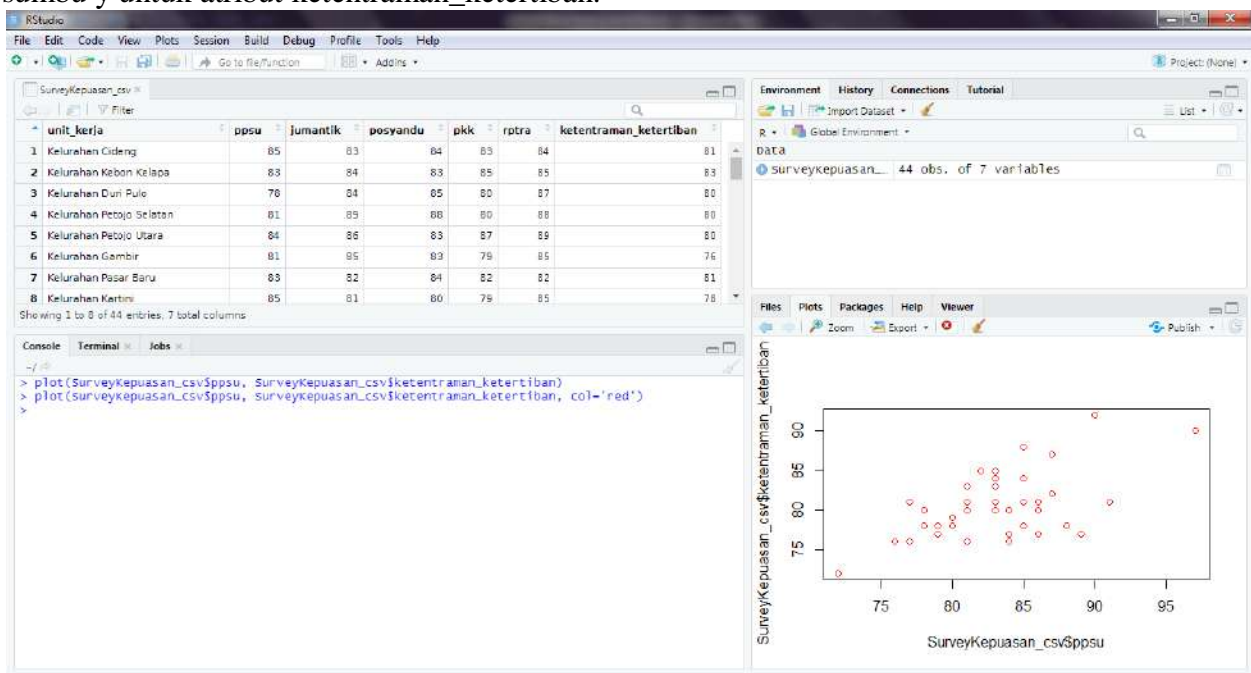
D. MODELLING

Data Modelling adalah proses yang digunakan untuk mendefinisikan dan menganalisis kebutuhan data yang diperlukan untuk mengetahui informasi yang dibutuhkan. Pada project ini saya akan melakukan modelling pada data SurveyKepuasan\_csv yang sebelumnya telah dilakukan preprocessing data. Terdapat 2 modelling yang akan dilakukan yaitu Regresi dan Clustering.

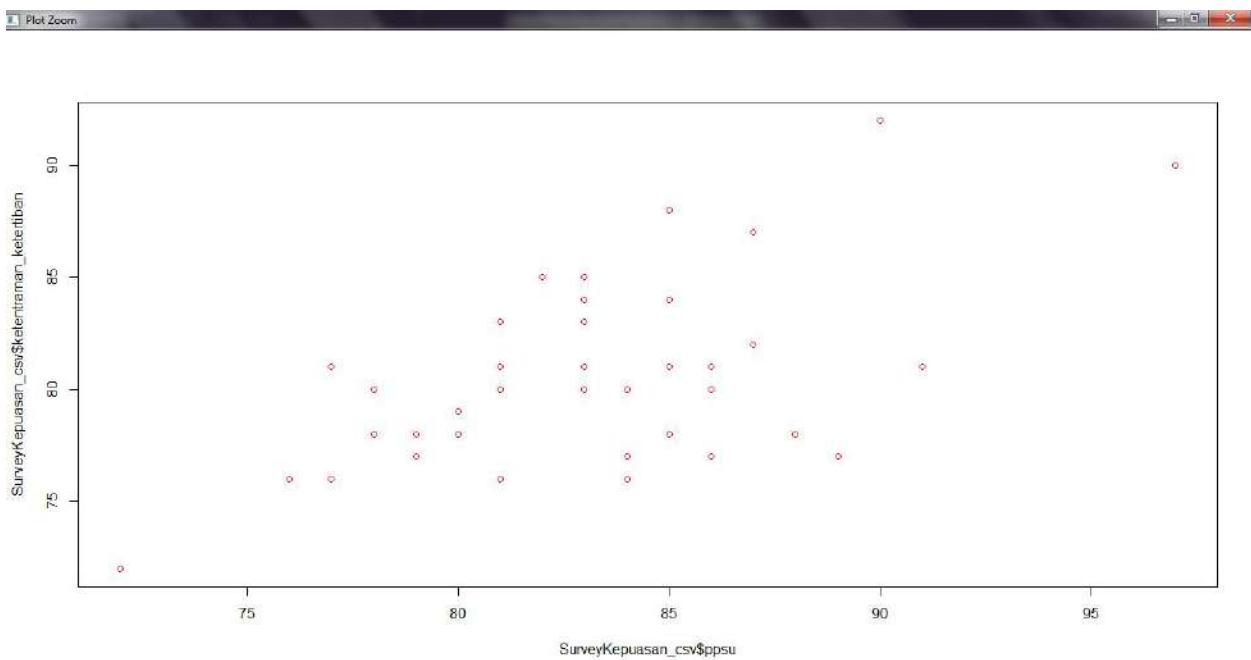
1. REGRESI

Regresi adalah suatu teknik pemodelan data untuk mengetahui hubungan dua atau lebih atribut atau untuk memprediksi nilai dari satu atribut dengan atribut lain. Atribut yang akan di regresi yaitu ppsu dan ketentraman\_ketertiban. Kedua atribut ini menghasilkan **korelasi positif**. hubungannya yaitu apabila index kepuasan pelayanan ppsu meningkat maka index ketentraman\_ketertiban pada tiap kelurahan di Jakarta Pusat juga akan meningkat. Langkah untuk melakukan regresi pada R Studio yaitu sebagai berikut :

Gunakan data yang telah dilakukan preprocessing data. Kemudian lakukan plot pada kedua atribut yaitu ppsu dan ketentraman\_ketertiban, dengan mengetikkan kode :  
plot(SurveyKepuasan\_csv\$ppsu, SurveyKepuasan\_csv\$ketentraman\_ketertiban, col='red'), Lalu akan muncul plot seperti tampilan di bawah. sumbu x untuk atribut ppsu dan sumbu y untuk atribut ketentraman\_ketertiban.



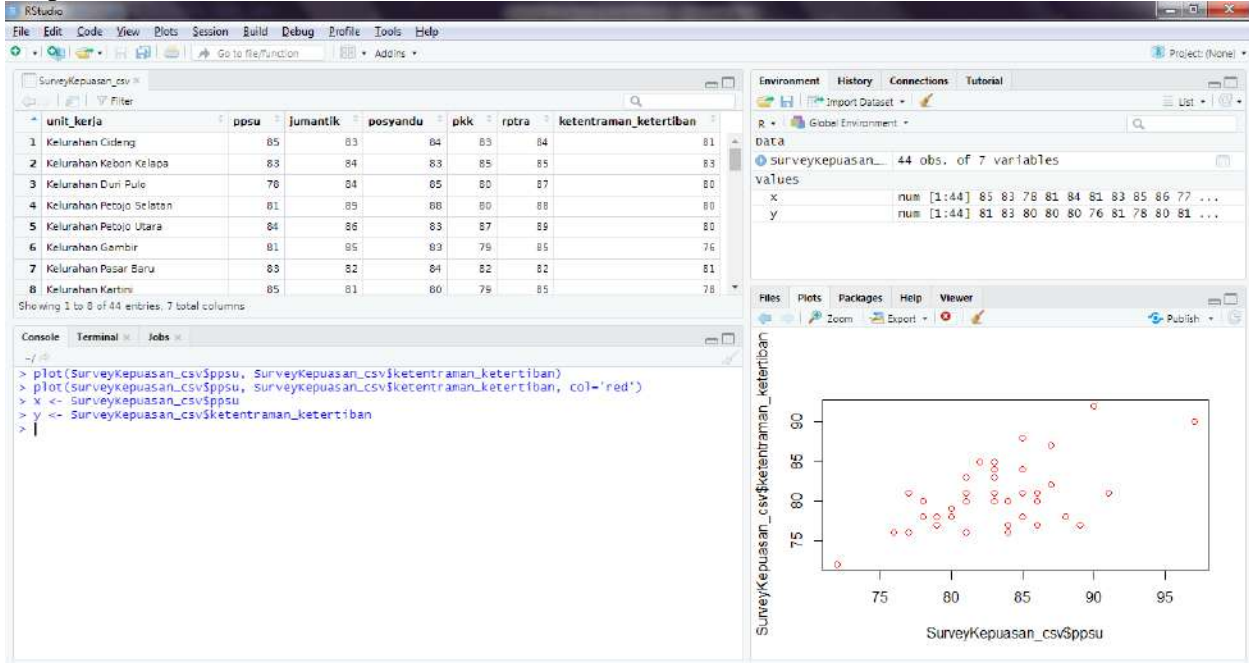
Tampilan Hasil plot dari atribut ppsu dan ketentraman\_ketertiban



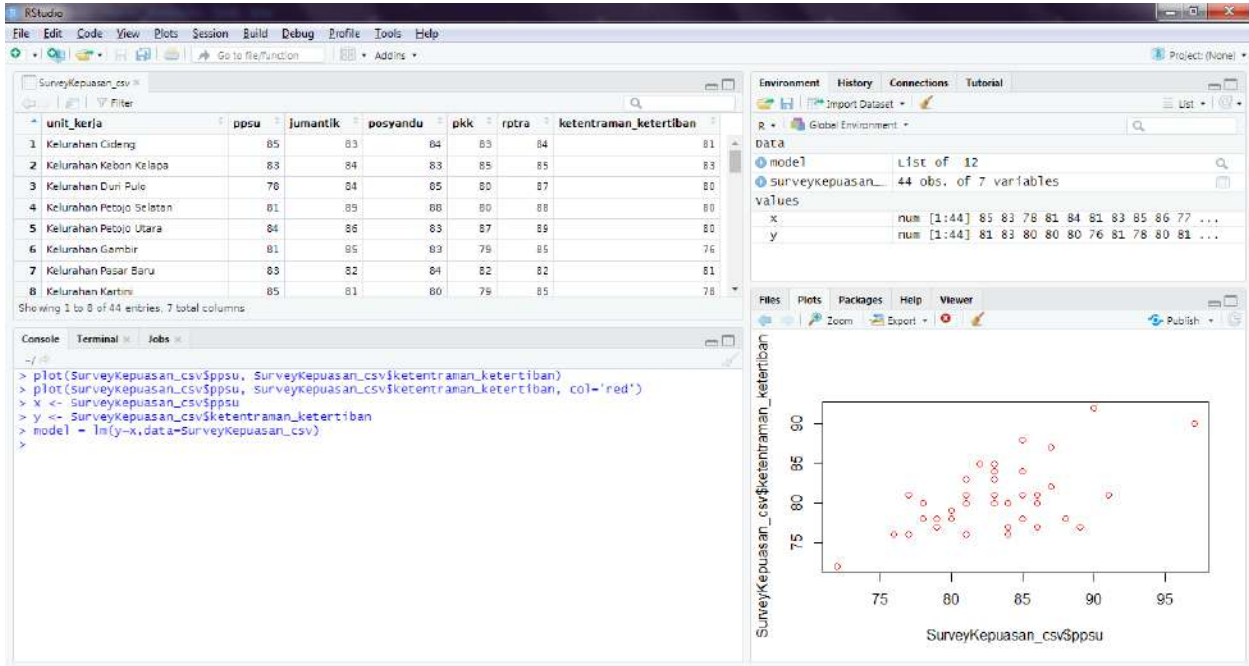
Selanjutnya lakukan prediksi masing-masing atribut dengan mengetikkan kode :

```
x <- SurveyKepuasan_csv$ppsu
y <- SurveyKepuasan_csv$ketentraman_ketertiban
```

Maka values x dan y akan muncul pada Global Environment dengan data berbentuk array seperti tampilan di bawah.

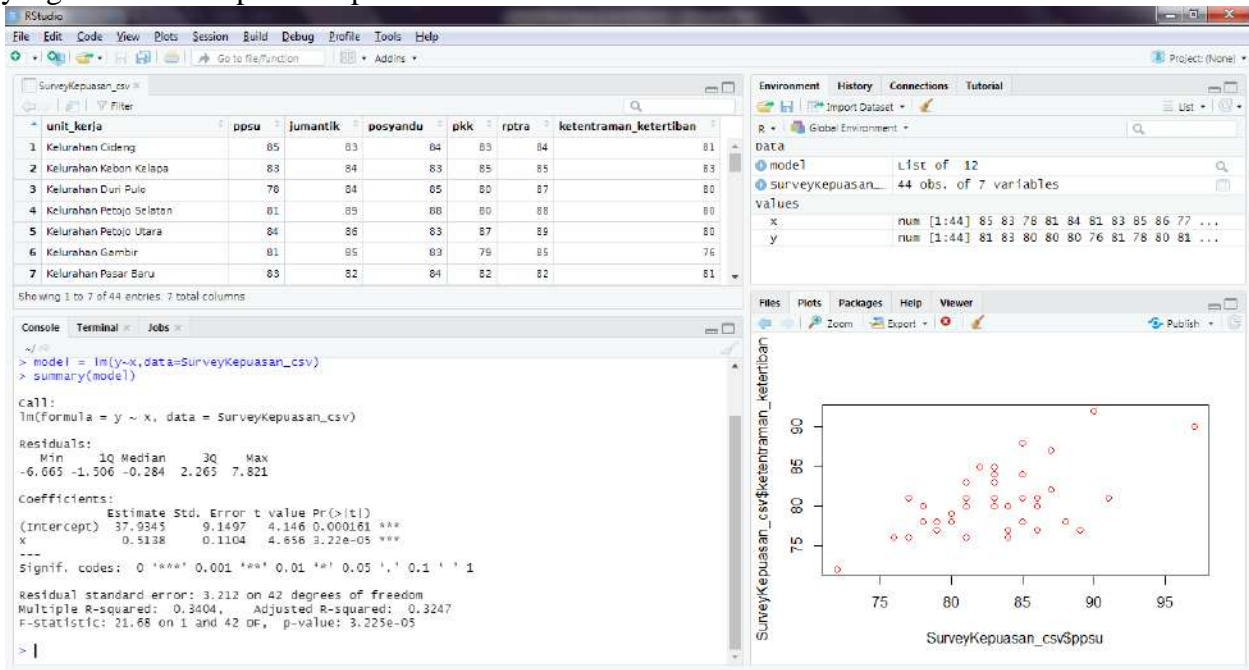


Lakukan model dengan membuat fungsi lm untuk memetakan y dan x menggunakan kode `model = lm(y~x,data=SurveyKepuasan_csv)`. Maka model akan muncul pada Data di Global Environment.

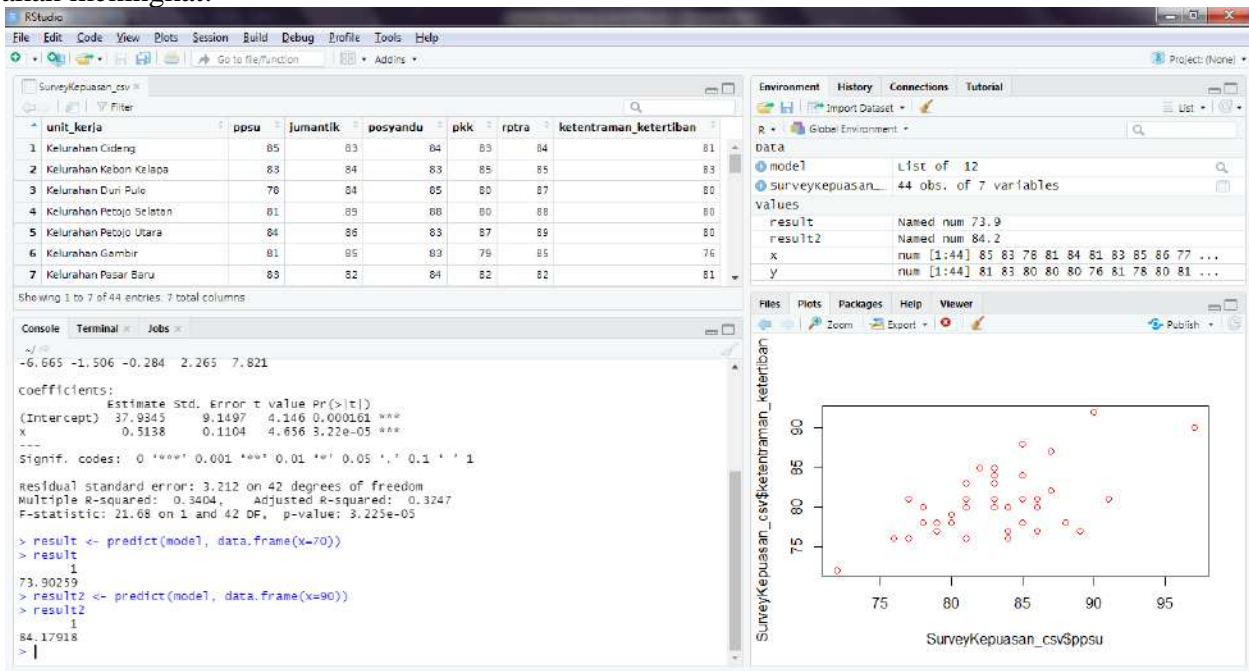




Kemudian ketikkan kode `summary(model)` untuk mengetahui nilai-nilai dari model . Output yang dihasilkan seperti tampilan di bawah.



Setelah itu lakukan prediksi kedua atribut. Dengan kode seperti di bawah. Disini saya melakukan prediksi dengan nilai `ppsu=x=70` lalu menghasilkan `73.90259` dan `x=90` menghasilkan `84.17918`. Dari hasil tersebut dapat diketahui bahwa kedua atribut menghasilkan korelasi positif , dimana apabila nilai index `ppsu` menurun maka index `ketentraman_ketertiban` juga akan menurun. Sebaliknya apabila nilai index `ppsu` melakukan kenaikan maka nilai `ketentraman_ketertiban` juga akan meningkat.

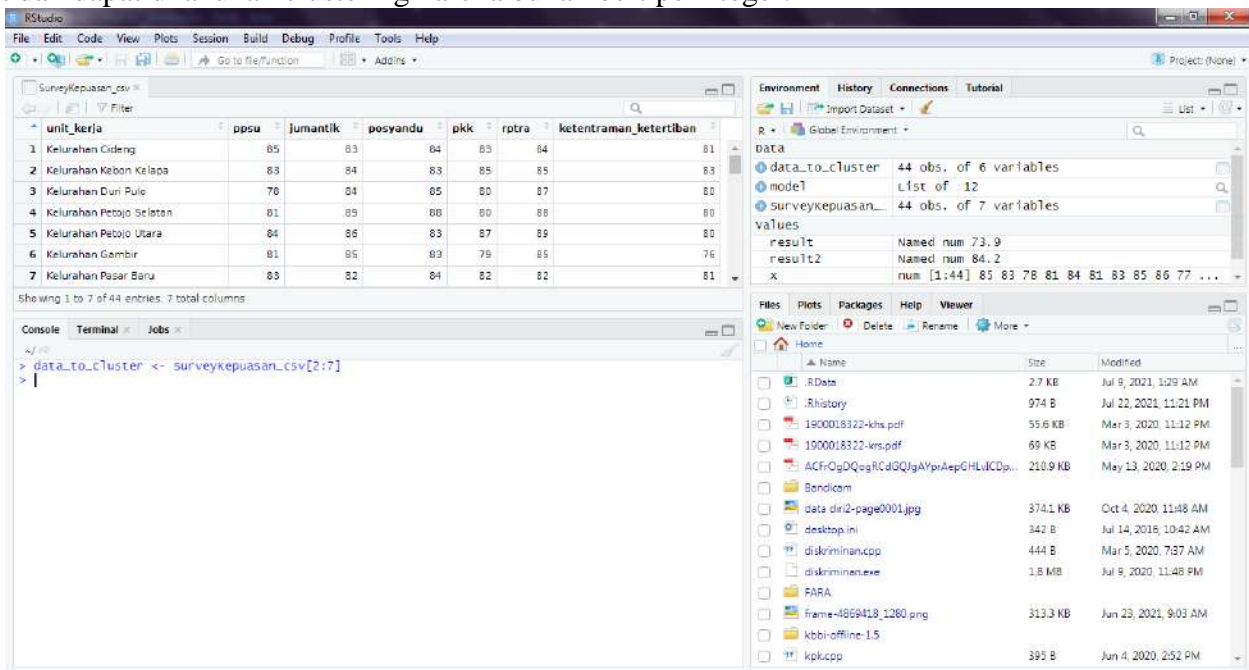


2. CLUSTERING

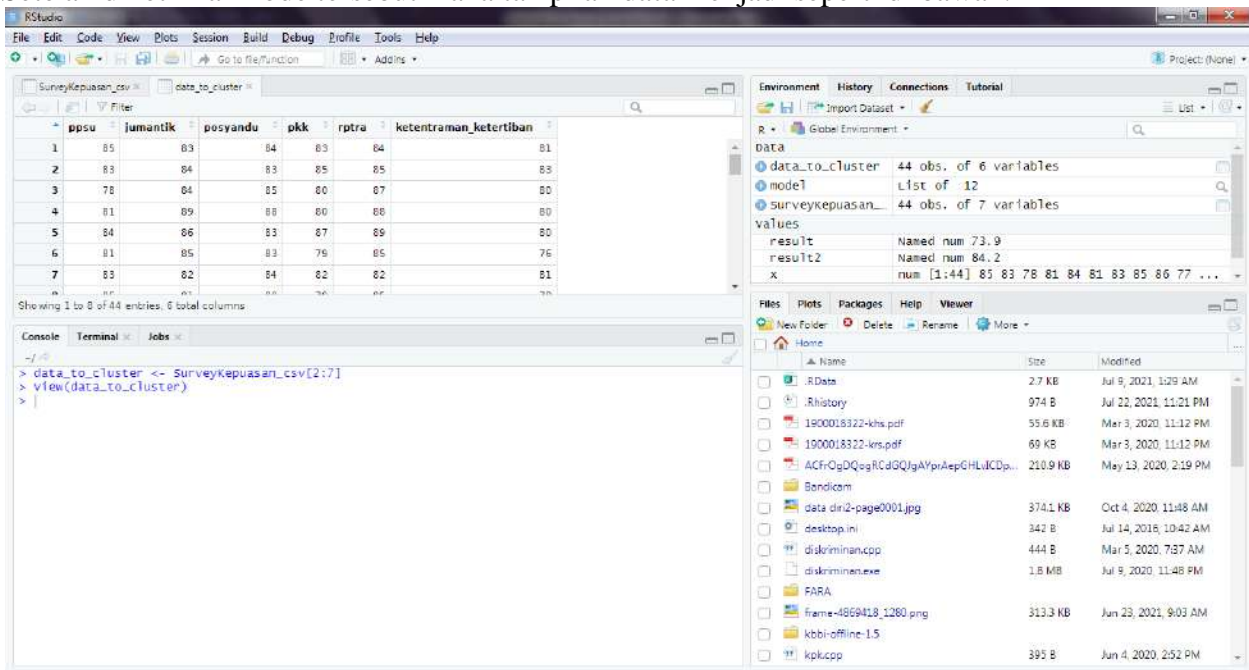
Clustering adalah metode penganalisaan data untuk mengelompokkan data dengan karakteristik data dengan membagi menjadi beberapa kluster. Disini saya akan melakukan clustering pada atribut yang bertipe data integer yaitu `ppsu`,`jumantik`, `posyandu`, `pkk`, `rptra`, dan `ketentraman_ketertiban`.



Langkah pertama untuk clustering adalah dengan mengetikkan kode `data_to_cluster <- SurveyKepuasan_csv[2:7]`. Maksudnya adalah untuk mengambil atribut yang akan dilakukan clustering yakni atribut 2 sampai 7 karena bertipe integer. Sedangkan atribut 1 yaitu `unit_kerja` tidak dapat dilakukan clustering karena bukan bertipe integer.



Setelah diketikkan kode tersebut maka tampilan data menjadi seperti di bawah.



Tampilan data secara keseluruhan setelah dilakukan pemilihan atribut integer.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

SurveyKepuasan\_csv data\_to\_cluster

Filter

	ppsu	jumantik	posyandu	pkk	rptra	ketentraran_ketertiban
1	85	83	84	83	84	81
2	83	84	83	85	85	83
3	78	84	85	80	87	80
4	81	89	88	80	88	80
5	84	86	83	87	89	80
6	81	85	83	79	85	76
7	83	82	84	82	82	81
8	85	81	80	79	85	78
9	86	85	82	84	86	80
10	77	87	77	80	76	81
11	79	81	81	80	76	78
12	81	83	81	79	85	80
13	82	86	82	81	86	85
14	80	78	80	77	77	78
15	77	75	74	75	85	76
16	83	81	90	82	82	80
17	83	80	80	81	81	80
18	76	76	76	76	76	76
19	78	78	78	78	85	78
20	91	80	86	85	88	81
21	73	76	79	78	80	77
22	80	80	80	79	85	79
23	83	84	78	84	84	83

Showing 1 to 23 of 44 entries, 6 total columns

Console

Environment History Connections Tutorial

Global Environment

data

data\_to\_cluster 44 obs. of 6 variables

model List of 12

surveyKepuasan\_csv 44 obs. of 7 variables

values

result Named num 73.9

result2 Named num 84.2

x num [1:44] 85 83 78 81 84 81 83 85 86 77 ...

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home

Name	Size	Modified
.RData	2.7 KB	Jul 9, 2021, 1:29 AM
.Rhistory	974 B	Jul 22, 2021, 11:21 PM
1900018322-4hs.pdf	55.6 KB	Mar 3, 2020, 11:12 PM
1900018322-lrs.pdf	69 KB	Mar 3, 2020, 11:12 PM
ACFrOgDQogRCdQigAYprAepGHLvCDpGRvdSYyHESRy9WFI81nPR9...	210.9 KB	May 13, 2020, 2:19 PM
Bandicam		
data_din2-page0001.jpg	374.1 KB	Oct 4, 2020, 11:48 AM
desktop.ini	342 B	Jul 14, 2018, 10:42 AM
diskriminan.cpp	444 B	Mar 5, 2020, 7:57 AM
diskriminan.exe	1.8 MB	Jul 9, 2020, 11:48 PM
FARA		
frame-4669418_1280.png	313.3 KB	Jun 23, 2021, 9:03 AM
hbbi-offline-1.5		
lplk.cpp	395 B	Jun 4, 2020, 2:52 PM

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

SurveyKepuasan\_csv data\_to\_cluster

Filter

	ppsu	jumantik	posyandu	pkk	rptra	ketentraran_ketertiban
23	87	84	78	86	85	82
24	86	85	81	82	85	81
25	81	81	82	83	85	81
26	72	75	82	76	85	72
27	80	80	80	80	85	79
28	79	78	78	78	85	78
29	85	85	83	83	85	84
30	84	86	84	82	87	76
31	87	92	96	92	90	80
32	81	77	75	79	91	83
33	83	83	82	83	85	83
34	86	87	83	86	78	77
35	87	84	83	83	85	87
36	98	91	86	89	92	92
37	85	84	82	78	85	88
38	88	89	88	88	89	78
39	89	84	83	82	86	77
40	83	84	85	86	87	85
41	84	86	83	82	87	77
42	81	83	84	84	86	80
43	88	78	82	78	88	79
44	89	82	83	83	86	84

Showing 22 to 44 of 44 entries, 6 total columns

Console

Environment History Connections Tutorial

Global Environment

data

data\_to\_cluster 44 obs. of 6 variables

model List of 12

surveyKepuasan\_csv 44 obs. of 7 variables

values

result Named num 73.9

result2 Named num 84.2

x num [1:44] 85 83 78 81 84 81 83 85 86 77 ...

Files Plots Packages Help Viewer

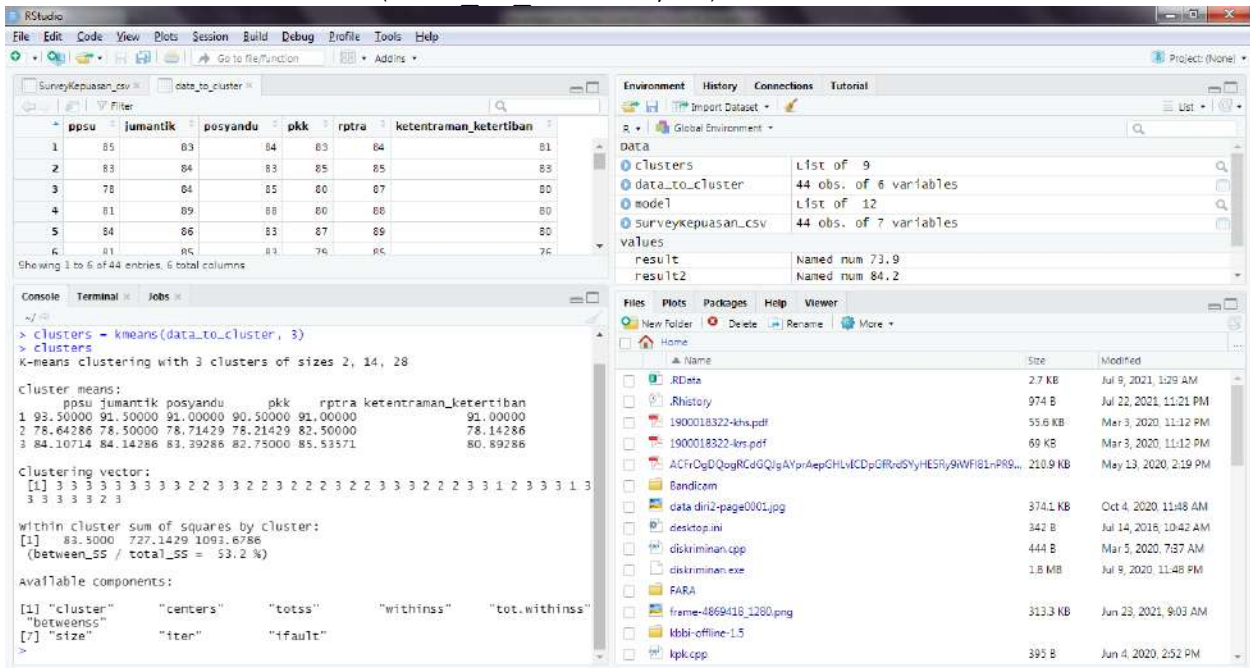
New Folder Delete Rename More

Home

Name	Size	Modified
.RData	2.7 KB	Jul 9, 2021, 1:29 AM
.Rhistory	974 B	Jul 22, 2021, 11:21 PM
1900018322-4hs.pdf	55.6 KB	Mar 3, 2020, 11:12 PM
1900018322-lrs.pdf	69 KB	Mar 3, 2020, 11:12 PM
ACFrOgDQogRCdQigAYprAepGHLvCDpGRvdSYyHESRy9WFI81nPR9...	210.9 KB	May 13, 2020, 2:19 PM
Bandicam		
data_din2-page0001.jpg	374.1 KB	Oct 4, 2020, 11:48 AM
desktop.ini	342 B	Jul 14, 2018, 10:42 AM
diskriminan.cpp	444 B	Mar 5, 2020, 7:57 AM
diskriminan.exe	1.8 MB	Jul 9, 2020, 11:48 PM
FARA		
frame-4669418_1280.png	313.3 KB	Jun 23, 2021, 9:03 AM
hbbi-offline-1.5		
lplk.cpp	395 B	Jun 4, 2020, 2:52 PM

Kemudian lakukan clustering dengan k-means:

- Clustering dengan nilai k=3  
clusters = kmeans(data to cluster, 3)



- Clustering dengan nilai k=5  
clusters = kmeans(data to cluster, 5)

