

# Tugas 2 Praktikum DS IF-A

Rifqi Maulana\_123200128

2022-11-06

```
knitr::opts_chunk$set(echo = TRUE)
```

## Tugas 2

### Instruksi

1. Kerjakan soal-soal yang ada! Jangan lupa tulis NAMA dan NIM pada author!
2. Jawablah dengan mengisi chunk dibawah soal!
3. Export hasil pekerjaan dalam format PDF/Word
4. Simpan file dengan format : nim\_Tugas2\_A ex : 123180062\_Tugas2\_A.pdf
5. Upload file .Rmd, .pdf dan gambar yg disimpan ke GDrive dan kumpulkan link yg dapat diakses ke SPADA

##Soal 1. Load library apa saja yang kira-kira digunakan! **2 poin**

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## v purrr   0.3.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(vroom)
library(here)
```

```
## Warning: package 'here' was built under R version 4.2.2
```

```
## here() starts at F:/semester 5/2. Course/Praktikum Data Science IF-E/Pertemuan 8 (Data Wraggling)/Tu
```

2. Load dataset “Data Jumlah Penduduk Berdasarkan Kewarganegaraan dan Jenis Kelamin 2020” dan tampilkan struktur dataset tersebut! **4 poin**

```
data_penduduk <- vroom(here("data-jumlah-penduduk-berdasarkan-kewarganegaraan-dan-jenis-kelamin-tahun-2020"))
```

```
## Rows: 1068 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (6): nama_provinsi, nama_kabupaten/kota, nama_kecamatan, nama_kelurahan,...
## dbl (2): tahun, jumlah
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
data_penduduk
```

```
## # A tibble: 1,068 x 8
##   tahun nama_provinsi nama_kabupaten/k~1 nama_~2 nama_~3 kewar~4 jenis~5 jumlah
##   <dbl> <chr>          <chr>          <chr>    <chr>    <chr>    <chr>    <dbl>
## 1 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. PAN~ WNI     Laki-L~ 3632
## 2 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. KEL~ WNI     Laki-L~ 3621
## 3 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. HAR~ WNI     Laki-L~ 1320
## 4 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. UNT~ WNI     Laki-L~ 1309
## 5 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. TID~ WNI     Laki-L~ 3062
## 6 2020 DKI Jakarta   KAB.ADM.KEP.SERIBU KEP. S~ P. PARI WNI     Laki-L~ 1871
## 7 2020 DKI Jakarta   JAKARTA PUSAT      GAMBIR    GAMBIR    WNI     Laki-L~ 1560
## 8 2020 DKI Jakarta   JAKARTA PUSAT      GAMBIR    CIDENG    WNI     Laki-L~ 9511
## 9 2020 DKI Jakarta   JAKARTA PUSAT      GAMBIR    PETOJO~ WNI     Laki-L~ 11238
## 10 2020 DKI Jakarta   JAKARTA PUSAT      GAMBIR    PETOJO~ WNI     Laki-L~ 9698
## # ... with 1,058 more rows, and abbreviated variable names
## #   1: 'nama_kabupaten/kota', 2: nama_kecamatan, 3: nama_kelurahan,
## #   4: kewarganegaraan, 5: jenis_kelamin
```

3. Hapus kolom tahun dan nama provinsi! **5 poin**

```
data_penduduk2 = data_penduduk[,-1:-2]
data_penduduk2
```

```
## # A tibble: 1,068 x 6
##   'nama_kabupaten/kota' nama_kecamatan nama_kelurahan kewarga~1 jenis~2 jumlah
##   <chr>                <chr>          <chr>          <chr>    <chr>    <dbl>
## 1 KAB.ADM.KEP.SERIBU   KEP. SERIBU UTR P. PANGGANG    WNI     Laki-L~ 3632
## 2 KAB.ADM.KEP.SERIBU   KEP. SERIBU UTR P. KELAPA      WNI     Laki-L~ 3621
## 3 KAB.ADM.KEP.SERIBU   KEP. SERIBU UTR P. HARAPAN    WNI     Laki-L~ 1320
## 4 KAB.ADM.KEP.SERIBU   KEP. SERIBU SLT P. UNTUNG JAWA WNI     Laki-L~ 1309
## 5 KAB.ADM.KEP.SERIBU   KEP. SERIBU SLT P. TIDUNG      WNI     Laki-L~ 3062
## 6 KAB.ADM.KEP.SERIBU   KEP. SERIBU SLT P. PARI        WNI     Laki-L~ 1871
## 7 JAKARTA PUSAT        GAMBIR          GAMBIR          WNI     Laki-L~ 1560
## 8 JAKARTA PUSAT        GAMBIR          CIDENG          WNI     Laki-L~ 9511
## 9 JAKARTA PUSAT        GAMBIR          PETOJO UTARA    WNI     Laki-L~ 11238
## 10 JAKARTA PUSAT       GAMBIR          PETOJO SELATAN WNI     Laki-L~ 9698
## # ... with 1,058 more rows, and abbreviated variable names 1: kewarganegaraan,
## #   2: jenis_kelamin
```

4. Pisahkan data WNI laki-laki dan perempuan! Gabungkan kedua data tersebut berdasarkan kelurahannya! **15 poin**

```
wni_laki<-filter(data_penduduk2, jenis_kelamin == 'Laki-Laki')
wni_perempuan<-filter(data_penduduk2, jenis_kelamin == 'Perempuan')
data_gabungan<-full_join(wni_laki,wni_perempuan, by ="nama_kelurahan")
data_gabungan
```

```
## # A tibble: 1,068 x 11
##   nama_kabupa~1 nama_~2 nama_~3 kewar~4 jenis~5 jumla~6 nama_~7 nama_~8 kewar~9
##   <chr>         <chr>    <chr>  <chr>    <chr>      <dbl> <chr>    <chr>    <chr>
## 1 KAB.ADM.KEP.~ KEP. S~ P. PAN~ WNI     Laki-L~   3632 KAB.AD~ KEP. S~ WNI
## 2 KAB.ADM.KEP.~ KEP. S~ P. PAN~ WNI     Laki-L~   3632 KAB.AD~ KEP. S~ WNA
## 3 KAB.ADM.KEP.~ KEP. S~ P. KEL~ WNI     Laki-L~   3621 KAB.AD~ KEP. S~ WNI
## 4 KAB.ADM.KEP.~ KEP. S~ P. KEL~ WNI     Laki-L~   3621 KAB.AD~ KEP. S~ WNA
## 5 KAB.ADM.KEP.~ KEP. S~ P. HAR~ WNI     Laki-L~   1320 KAB.AD~ KEP. S~ WNI
## 6 KAB.ADM.KEP.~ KEP. S~ P. HAR~ WNI     Laki-L~   1320 KAB.AD~ KEP. S~ WNA
## 7 KAB.ADM.KEP.~ KEP. S~ P. UNT~ WNI     Laki-L~   1309 KAB.AD~ KEP. S~ WNI
## 8 KAB.ADM.KEP.~ KEP. S~ P. UNT~ WNI     Laki-L~   1309 KAB.AD~ KEP. S~ WNA
## 9 KAB.ADM.KEP.~ KEP. S~ P. TID~ WNI     Laki-L~   3062 KAB.AD~ KEP. S~ WNI
## 10 KAB.ADM.KEP.~ KEP. S~ P. TID~ WNI     Laki-L~   3062 KAB.AD~ KEP. S~ WNA
## # ... with 1,058 more rows, 2 more variables: jenis_kelamin.y <chr>,
## #   jumlah.y <dbl>, and abbreviated variable names 1: 'nama_kabupaten/kota.x',
## #   2: nama_kecamatan.x, 3: nama_kelurahan, 4: kewarganegaraan.x,
## #   5: jenis_kelamin.x, 6: jumlah.x, 7: 'nama_kabupaten/kota.y',
## #   8: nama_kecamatan.y, 9: kewarganegaraan.y
```

5. Rapikan data gabungan sehingga hanya ada 1 kolom kabupate/kota, kecamatan, kelurahan, kewarganegaraan, serta 2 kolom jumlah penduduk berdasarkan jenis kelamin! Ubah nama kolom menjadi : kabupaten\_kota, kecamatan, kelurahan, kewarganegaraan, jumlah\_laki, jumlah\_perempuan sehingga lebih mudah dibaca! **20 poin**

```
new_data <-data_gabungan%>%
  select('nama_kabupaten/kota.x','nama_kecamatan.x','nama_kelurahan','kewarganegaraan.x',
        'jenis_kelamin.x','jumlah.x','jenis_kelamin.y','jumlah.y')
new_data <- new_data %>%
  rename('kabupaten_kota' = 'nama_kabupaten/kota.x', 'kecamatan'='nama_kecamatan.x', 'kelurahan'='nama_~8',
        'jumlah_laki'='jumlah.x', 'jumlah_perempuan' = 'jumlah.y')
new_data
```

```
## # A tibble: 1,068 x 8
##   kabupaten_kota kecamatan kelur~1 kewar~2 jenis~3 jumla~4 jenis~5 jumla~6
##   <chr>          <chr>    <chr>  <chr>    <chr>      <dbl> <chr>      <dbl>
## 1 KAB.ADM.KEP.SERIBU KEP. SERI~ P. PAN~ WNI     Laki-L~   3632 Peremp~   3474
## 2 KAB.ADM.KEP.SERIBU KEP. SERI~ P. PAN~ WNI     Laki-L~   3632 Peremp~     0
## 3 KAB.ADM.KEP.SERIBU KEP. SERI~ P. KEL~ WNI     Laki-L~   3621 Peremp~   3550
## 4 KAB.ADM.KEP.SERIBU KEP. SERI~ P. KEL~ WNI     Laki-L~   3621 Peremp~     0
## 5 KAB.ADM.KEP.SERIBU KEP. SERI~ P. HAR~ WNI     Laki-L~   1320 Peremp~   1328
## 6 KAB.ADM.KEP.SERIBU KEP. SERI~ P. HAR~ WNI     Laki-L~   1320 Peremp~     0
## 7 KAB.ADM.KEP.SERIBU KEP. SERI~ P. UNT~ WNI     Laki-L~   1309 Peremp~   1238
## 8 KAB.ADM.KEP.SERIBU KEP. SERI~ P. UNT~ WNI     Laki-L~   1309 Peremp~     0
## 9 KAB.ADM.KEP.SERIBU KEP. SERI~ P. TID~ WNI     Laki-L~   3062 Peremp~   3016
```

```
## 10 KAB.ADM.KEP.SERIBU KEP. SERI~ P. TID~ WNI      Laki-L~      3062 Peremp~      0
## # ... with 1,058 more rows, and abbreviated variable names 1: kelurahan,
## # 2: kewarganegaraan.x, 3: jenis_kelamin.x, 4: jumlah_laki,
## # 5: jenis_kelamin.y, 6: jumlah_perempuan
```

6. Kelompokkan data tersebut berdasarkan kecamatan dan lakukan penjumlahan sehingga jumlah penduduk laki-laki dan perempuan tergabung untuk masing-masing kecamatan! Ket kolom : kecamatan, total\_laki, total\_perempuan **10 poin**

```
total_laki = aggregate(new_data$jumlah_laki,list(new_data$kecamatan), FUN = sum)
total_perempuan = aggregate(new_data$jumlah_perempuan,list(new_data$kecamatan), FUN = sum)
total_kecamatan = inner_join(total_laki,total_perempuan, by = "Group.1")
total_kecamatan = rename(total_kecamatan, 'kecamatan'='Group.1','total_laki' = 'x.x', 'total_perempuan'='y.y')
total_kecamatan
```

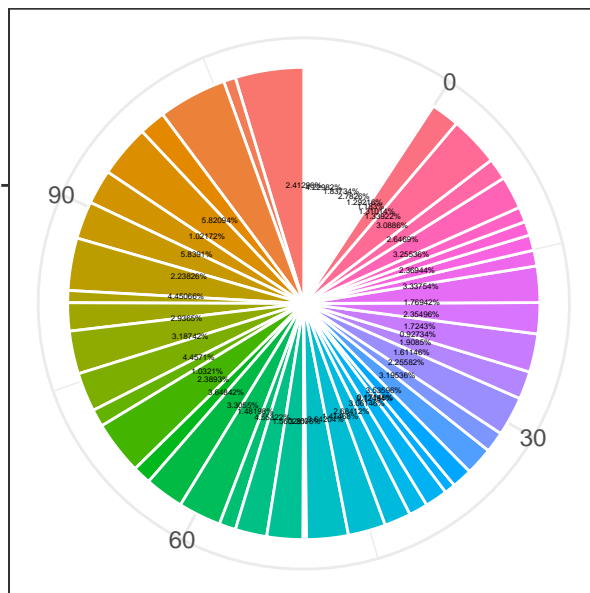
```
##      kecamatan total_laki total_perempuan
## 1      CAKUNG      582094      563964
## 2  CEMPAKA PUTIH      102172      102484
## 3    CENGKARENG      583910      568730
## 4      CILANDAK      223826      227422
## 5    CILINCING      445066      435428
## 6      CIPAYUNG      293650      289208
## 7      CIRACAS      318742      314556
## 8    DUREN SAWIT      445710      447930
## 9        GAMBIR      103210      101282
## 10 GROGOL PETAMBURAN      238930      241446
## 11      JAGAKARSA      364842      361510
## 12    JATINEGARA      330550      326140
## 13    JOHAR BARU      148198      144144
## 14    KALI DERES      455322      444674
## 15  KEBAYORAN BARU      156028      158094
## 16  KEBAYORAN LAMA      330760      328716
## 17    KEBON JERUK      364204      361852
## 18    KELAPA GADING      141468      146970
## 19    KEMAYORAN      268412      264566
## 20    KEMBANGAN      306146      303698
## 21  KEP. SERIBU SLT      12484      12126
## 22  KEP. SERIBU UTR      17146      16704
## 23        KOJA      353596      344038
## 24    KRAMAT JATI      319536      315852
## 25      MAKASAR      225582      223326
## 26  MAMPANG PRAPATAN      161146      157480
## 27      MATRAMAN      190850      191442
## 28      MENTENG      92734      91518
## 29    PADEMANGAN      172430      166734
## 30      PALMERAH      235496      230424
## 31    PANCORAN      176942      176664
## 32    PASAR MINGGU      333754      330204
## 33    PASAR REBO      236944      233234
## 34    PENJARINGAN      325536      318068
## 35    PESANGGRAHAN      264690      264758
## 36    PULO GADUNG      308860      310382
## 37    SAWAH BESAR      133922      133544
```

```
## 38          SENEN      131014      127768
## 39      SETIA BUDI    118300      117758
## 40      TAMAN SARI    129216      129642
## 41      TAMBORA      278260      268514
## 42      TANAH ABANG   183734      179554
## 43      TANJUNG PRIOK 422982      416608
## 44          TEBET     241296      243396
```

7. Buatlah diagram pie menggunakan data laki-laki pada nomor 6. seperti contoh yg diberikan pada gambar 1. **15 poin**

```
ypos = cumsum(total_kecamatan$total_laki/100000) - (1/2) * (total_kecamatan$total_laki/100000)
ypos = 100 - ypos
l <- ggplot() + theme_bw() +
  geom_bar(aes(x = "", y =total_kecamatan$total_laki/100000, fill = total_kecamatan$kecamatan),
    width = 0.2, stat = "identity", color = "white") +
  coord_polar(theta = "y") +
  theme(plot.title = element_text(hjust = 0.5, size =10), legend.key.size = unit(0.2, "cm")) +
  theme(legend.text = element_text(size = 5, colour = "black"))+
  labs(title="Diagram Pie Penduduk Laki-Laki di Jakarta",
    x=" ",
    y= "Total Laki-Laki (satuan jutaan)")+
  guides(fill=guide_legend(title="Daftar Kecamatan")) +
  geom_text(aes(x = "", y =ypos, label = paste0(total_kecamatan$total_laki/100000, "%")), size = 1)
1
```

Diagram Pie Penduduk Laki-Laki di Jakarta



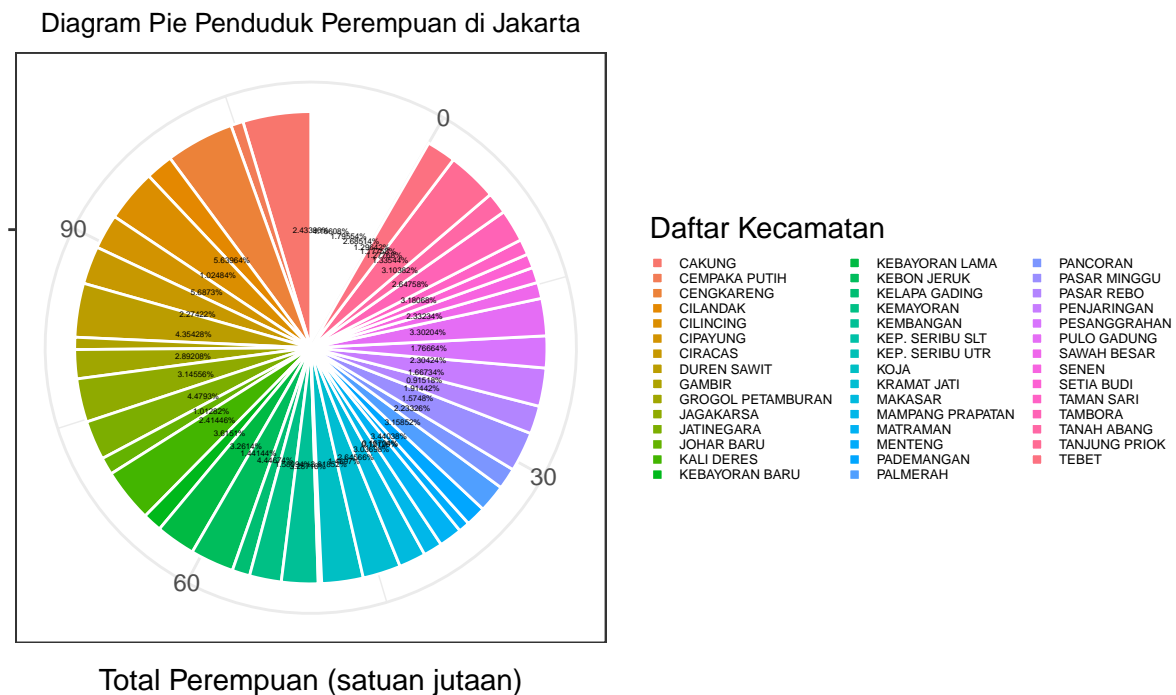
Daftar Kecamatan

CAKUNG	KEBAYORAN LAMA	PANCORAN
CEMPAKA PUTIH	KEBON JERUK	PASAR MINGGU
CENGKARENG	KELAPA GADING	PASAR REBO
CILANDAK	KEMAYORAN	PENJARINGAN
CILINCING	KEMBANGAN	PESANGGRAHAN
CIPAYUNG	KEP. SERIBU SLT	PULO GADUNG
CIRACAS	KEP. SERIBU UTR	SAWAH BESAR
DUREN SAWIT	KOJA	SEKEN
GAMBIR	KRAMAT JATI	SETIA BUDI
GROGOL PETAMBURAN	MAKASAR	TAMAN SARI
JAGAKARSA	MAMPANG PRAPATAN	TAMBORA
JATINEGARA	MATRAMAN	TANAH ABANG
JOHAR BARU	MENTENG	TANJUNG PRIOK
KALI DERES	PADEMANGAN	TEBET
KEBAYORAN BARU	PALMERAH	

Total Laki-Laki (satuan jutaan)

8. Lakukan hal yang sama untuk data perempuan! seperti contoh gambar 2. **7 poin**

```
ypos = cumsum(total_kecamatan$total_perempuan/100000) - (1/2) * (total_kecamatan$total_perempuan/100000)
ypos = 100- ypos
p <- ggplot() + theme_bw() +
  geom_bar(aes(x = "", y =total_kecamatan$total_perempuan/100000, fill = total_kecamatan$kecamatan),
    width = 0.1, stat = "identity", color = "white") +
  coord_polar(theta = "y") +
  theme(plot.title = element_text(hjust = 0.5, size =10),legend.key.size = unit(0.2, "cm")) +
  theme(legend.text = element_text(size = 5, colour = "black"))+
  labs(title="Diagram Pie Penduduk Perempuan di Jakarta",
    x=" ",
    y= "Total Perempuan (satuan jutaan)")+
  guides(fill=guide_legend(title="Daftar Kecamatan")) +
  geom_text(aes(x = "", y = ypos, label = paste0(total_kecamatan$total_perempuan/100000, "%")), size = 10)
p
```



9. Dengan menggunakan dataset asli, ambil data kewarganegaraan WNI dan kelompokkan data berdasarkan kecamatan serta lakukan penjumlahan sehingga jumlah penduduk terkumpul berdasarkan kecamatan. **8 poin**

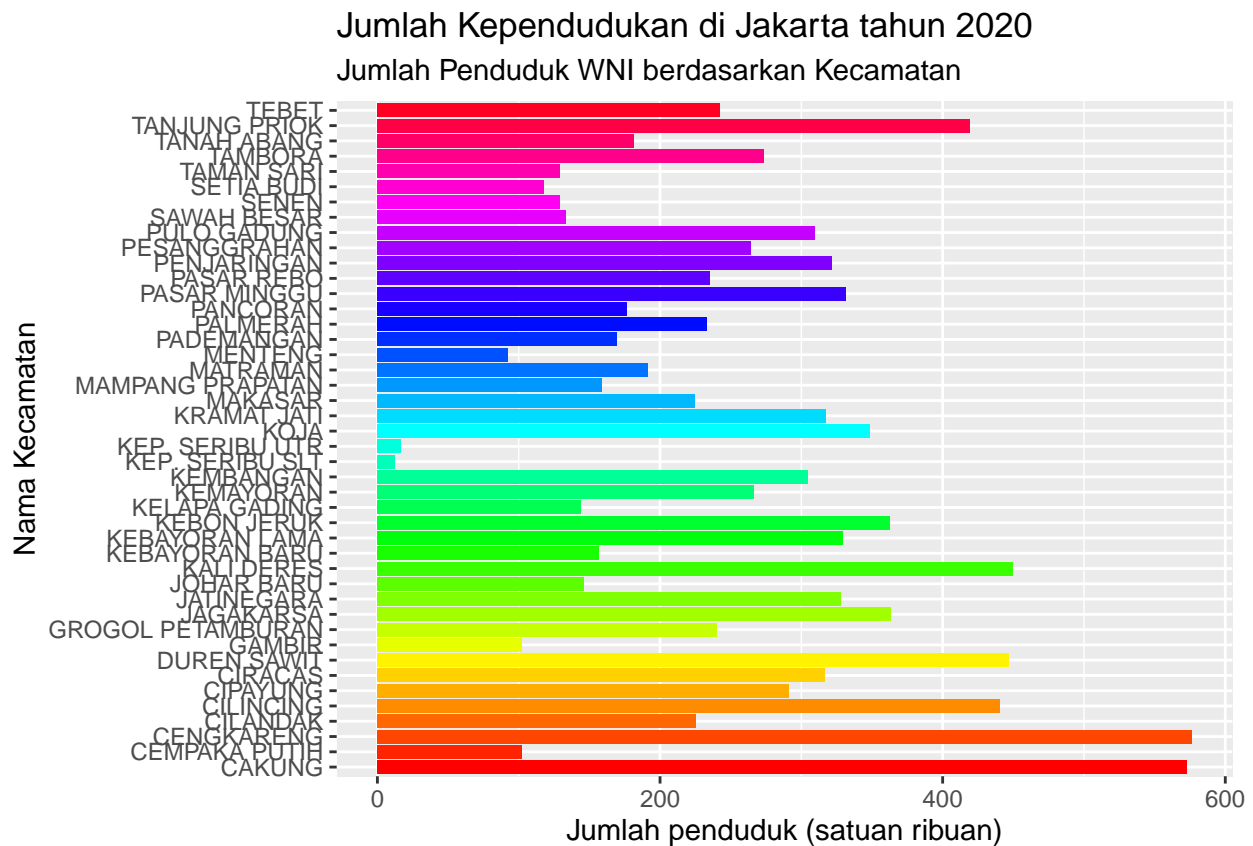
```
wni_kec = filter(data_penduduk, kewarganegaraan == 'WNI')
wni_kec = aggregate(wni_kec$jumlah,
```

```
list(wni_kec$nama_kecamatan), FUN = sum)
wni_kec
```

```
##           Group.1           x
## 1           CAKUNG 572996
## 2    CEMPAKA PUTIH 102273
## 3      CENGKARENG 576216
## 4      CILANDAK 225332
## 5      CILINCING 440235
## 6      CIPAYUNG 291416
## 7      CIRACAS 316639
## 8    DUREN SAWIT 446773
## 9           GAMBIR 102190
## 10 GROGOL PETAMBURAN 240015
## 11          JAGAKARSA 363091
## 12          JATINEGARA 328292
## 13          JOHAR BARU 146160
## 14          KALI DERES 449946
## 15    KEBAYORAN BARU 156737
## 16    KEBAYORAN LAMA 329442
## 17      KEBON JERUK 362899
## 18    KELAPA GADING 143976
## 19      KEMAYORAN 266309
## 20      KEMBANGAN 304815
## 21    KEP. SERIBU SLT 12305
## 22    KEP. SERIBU UTR 16925
## 23              KOJA 348804
## 24      KRAMAT JATI 317673
## 25          MAKASAR 224449
## 26    MAMPANG PRAPATAN 159079
## 27          MATRAMAN 191134
## 28          MENTENG 92066
## 29      PADEMANGAN 169427
## 30          PALMERAH 232874
## 31          PANCORAN 176719
## 32      PASAR MINGGU 331825
## 33      PASAR REBO 235073
## 34      PENJARINGAN 321540
## 35      PESANGGRAHAN 264667
## 36      PULO GADUNG 309558
## 37      SAWAH BESAR 133550
## 38          SENEN 129355
## 39      SETIA BUDI 117851
## 40      TAMAN SARI 129366
## 41          TAMBORA 273331
## 42      TANAH ABANG 181496
## 43    TANJUNG PRIOK 419555
## 44          TEBET 242259
```

10. Menggunakan data nomor 9, buatlah visualisasi data seperti contoh gambar 3! **10 poin**

```
k <- ggplot(wni_kec, aes(x/1000, Group.1)) +
  geom_bar( stat = "identity", fill = rainbow(n =length(wni_kec$Group.1)))
k <- k +
  labs(title = "Jumlah Kependudukan di Jakarta tahun 2020",
        subtitle = "Jumlah Penduduk WNI berdasarkan Kecamatan",
        y = "Nama Kecamatan",
        x= "Jumlah penduduk (satuan ribuan)")
k
```



11. Simpan seluruh visualisasi ke dalam bentuk gambar (jpg/png) **4 poin**

```
ggsave(filename = "l.jpeg", plot = l)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave(filename = "p.jpeg", plot = p)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave(filename = "k.jpeg", plot = k)
```

```
## Saving 6.5 x 4.5 in image
```