STK372 Komputasi Statistik 2
Pemrograman Grafik
Package ggplot2

Dr. Agus M Soleh

IPB University
— Bogor Indonesia —

agusms@apps.ipb.ac.id

Outline

Pengantar

Apa ggplot2?

- Implementasi *the Grammar of Graphics* oleh Leland Wilkinson
- Dibuat oleh Hadley Wickham (sewaktu masih mahasiswa Iowa State)
- Merupakan "sistem ketiga" grafik R (sebelumnya: `base` and `lattice`)
- Tersedia di CRAN via `install.packages()`
- Grammar of graphics merepresentasikan dan mengabstraksi ide atau objek grafik

Pengantar

dari buku ggplot2:

"In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system"

Komponen Grammar

Terdapat 4 komponen grammar:

- geom: geometric "bentuk" untuk menampilkan data
  `bar, point, line, ribbon, text, dll`
- aesthetic: atribut untuk mengatur bagaimana geom ditampilkan
  `x position, y position, color, fill, shape, size, dll`
- stat: transformasi data sebelum geom digunakan
  contoh: histogram menggunakan bin, dll
- scale: konversi data mentah (raw)

Fungsi untuk Grammar Graphics

Package `ggplot2`, menyediakan dua fungsi untuk membuat grammar graphics:

- `qplot()`:
  - Atau bisa disebut quick plot
  - Menggunakan sebagian konsep dari grammar graphics
  - Dirancang untuk lebih familiar dengan fungsi `plot()`
  - Membuat grafik lebih mudah dan cepat
- `ggplot()`:
  - Memberikan kapabilitas penuh dari grammar graphics

Fungsi qplot()

- Fungsi qplot() sangat mudah digunakan
- Hanya membutuhkan:
    - satu (atau dua) nama peubah data yang akan ditampilkan
    - data frame yang memuat peubah tersebut (optional)
- qplot() akan menginspeksi peubah tersebut dan mencoba membuat plot yang masuk akal

Ilustrasi qplot()

Ilustrasi menggunakan Data: mtcars (Motor Trend Car Road Tests)

```
Description:
The data was extracted from the 1974 Motor Trend US magazine,
    and comprises fuel consumption and 10 aspects of automobile
     design and performance for 32 automobiles (197374 models).

Format:
A data frame with 32 observations on 11 (numeric) variables.
[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp  Displacement (cu.in.)
[, 4] hp  Gross horsepower
[, 5] drat  Rear axle ratio
[, 6] wt  Weight (1000 lbs)
[, 7] qsec  1/4 mile time
[, 8] vs  Engine (0 = V-shaped, 1 = straight)
[, 9] am  Transmission (0 = automatic, 1 = manual)
[,10] gear  Number of forward gears
[,11] carb  Number of carburetors
```
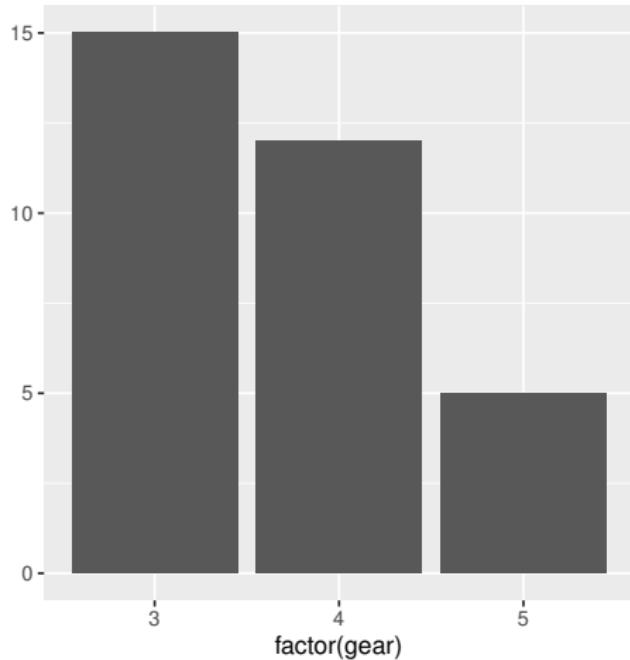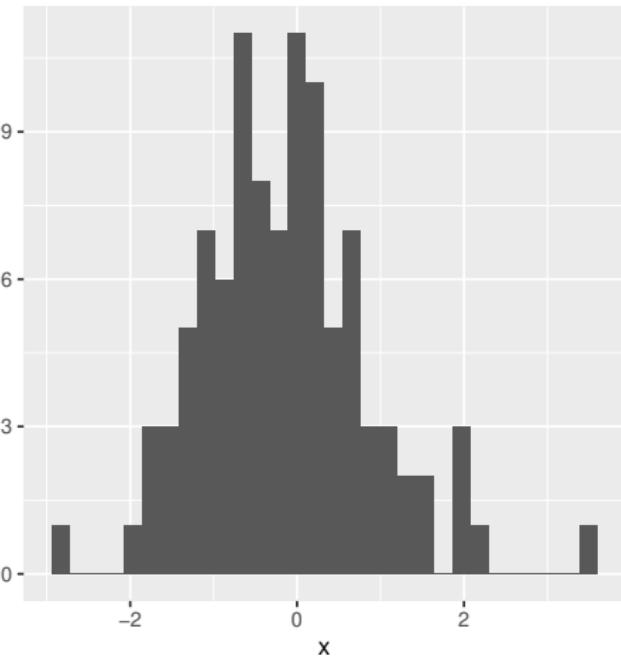
Ilustrasi `qplot()`: satu peubah kategorik

```
> qplot(factor(gear),data=mtcars)
```
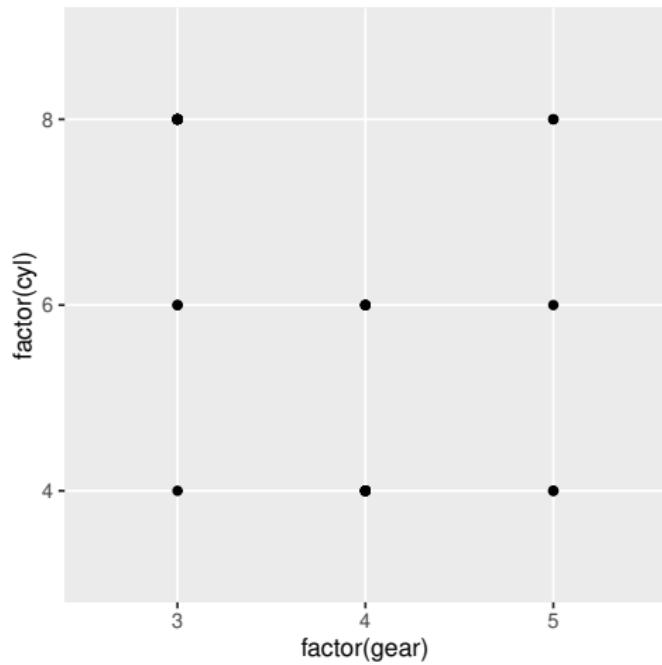
Ilustrasi `qplot()`: satu peubah numerik

```
> dtx <- data.frame(x=rnorm(100))
> qplot(x,data=dtx)
```
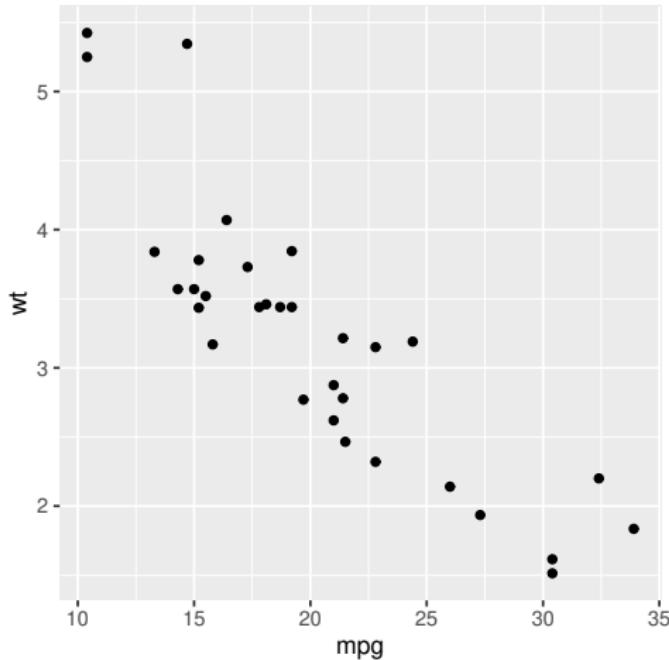
Ilustrasi qplot(): dua peubah kategorik

```
> qplot(factor(gear), factor(cyl), data=mtcars)   # tidak
bermanfaat
```
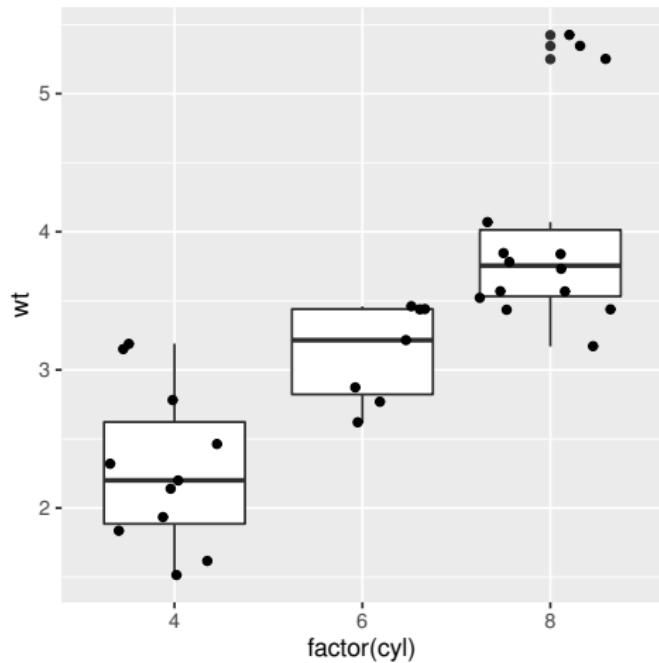
Ilustrasi `qplot()`: dua peubah numerik

```
> qplot(mpg, wt, data = mtcars)
```

Ilustrasi `qplot()`: dua peubah kategorik dan numerik

```
> qplot(factor(cyl), wt, data = mtcars, geom = c("boxplot", "
jitter"))
```

Penutup `qplot()`

- `qplot()` sudah sangat memadai, tetapi tidak memberikan kapabilitas penuh ggplot

Fungsi `ggplot()`
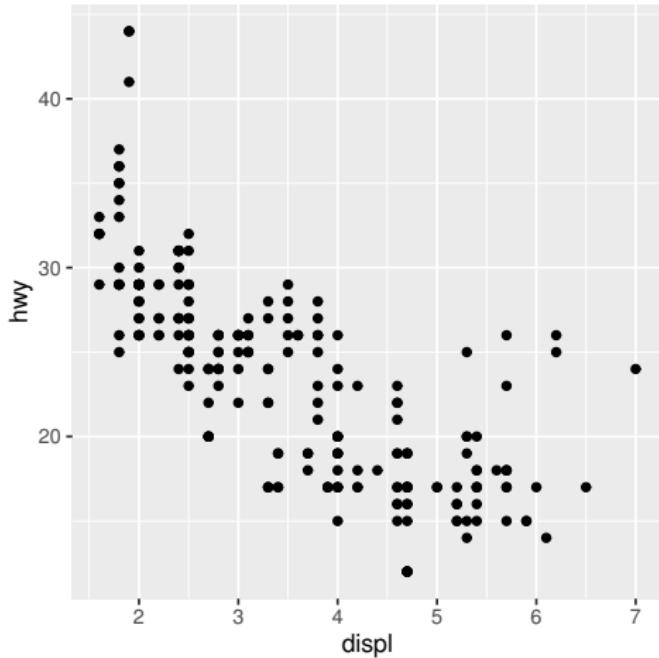
- Fungsi `ggplot()` memberikan objek ggplot
- Argumen dalam ggplot menetapkan data yang digunakan dan aesthetics
- Untuk menggambarkan grafik dibutuhkan layer
- Menambahkan layer dapat menggunakan fungsi `layer`, fungsi `stat`, atau fungsi `geom`

Ilustrasi

```
> ggplot(mpg, aes(displ, hwy
)) + geom_point()
```
adalah sama seperti:

```
> ggplot(mpg, aes(displ, hwy
    )) +
+      layer(geom = "point",
    stat = "identity",
    position = "identity",
+               params = list(na
    .rm = FALSE)
+      )
```

geom dan stat

1. Fungsi `geom` adalah layer untuk bentuk grafik, sedangkan fungsi `stat` adalah layer untuk data yang ditampilkan

2. Gunakan salah satunya

3. `geom` dan `stat` umumnya memiliki default berpasangan, tetapi ada beberapa geom dan stat yang defaultnya berbeda

   1. `geom_density` defaultnya adalah `stat_density`
   2. `stat_density` defaultnya dalah `geom_area`

4. Sehingga hati-hati dalam menggunakannya

Fungsi-fungsi geom

```
> apropos("^geom")
 [1] "Geom"              "geom_abline"       "geom_area"         "geom_bar"          "geom_bin2d"
 [6] "geom_blank"        "geom_boxplot"      "geom_col"          "geom_contour"      "geom_contour_filled"
[11] "geom_count"        "geom_crossbar"     "geom_curve"        "geom_density"      "geom_density_2d"
[16] "geom_density2d"    "geom_dotplot"      "geom_errorbar"     "geom_errorbarh"    "geom_freqpoly"
[21] "geom_hex"          "geom_histogram"    "geom_hline"        "geom_jitter"       "geom_label"
[26] "geom_line"         "geom_linerange"    "geom_map"          "geom_path"         "geom_point"
[31] "geom_pointrange"   "geom_polygon"      "geom_qq"           "geom_qq_line"      "geom_quantile"
[36] "geom_raster"       "geom_rect"         "geom_ribbon"       "geom_rug"          "geom_segment"
[41] "geom_sf"           "geom_sf_label"     "geom_sf_text"      "geom_smooth"       "geom_spoke"
[46] "geom_step"         "geom_text"         "geom_tile"         "geom_violin"       "geom_vline"
[51] "GeomAbline"        "GeomAnnotationMap" "GeomArea"          "GeomBar"           "GeomBlank"
[56] "GeomBoxplot"       "GeomCol"           "GeomContour"       "GeomCrossbar"      "GeomCurve"
[61] "GeomCustomAnn"     "GeomDensity"       "GeomDensity2d"     "GeomDotplot"       "GeomErrorbar"
[66] "GeomErrorbarh"     "GeomHex"           "GeomHline"         "GeomLabel"         "GeomLine"
[71] "GeomLinerange"     "GeomLogticks"      "GeomMap"           "GeomPath"          "GeomPoint"
[76] "GeomPointrange"    "GeomPolygon"       "GeomQuantile"      "GeomRaster"        "GeomRasterAnn"
[81] "GeomRect"          "GeomRibbon"        "GeomRug"           "GeomSegment"       "GeomSf"
[86] "GeomSmooth"        "GeomSpoke"         "GeomStep"          "GeomText"          "GeomTile"
[91] "GeomViolin"        "GeomVline"
```

Fungsi-fungsi stat

```
> apropos("^stat")
 [1] "stat"                "Stat"                "stat.anova"          "stat_bin"             "stat_bin_2d"
 [6] "stat_bin_hex"        "stat_bin2d"          "stat_binhex"         "stat_boxplot"         "stat_contour"
[11] "stat_contour_filled" "stat_count"          "stat_density"        "stat_density_2d"      "stat_density2d"
[16] "stat_ecdf"           "stat_ellipse"        "stat_function"       "stat_identity"        "stat_qq"
[21] "stat_qq_line"        "stat_quantile"       "stat_sf"             "stat_sf_coordinates"  "stat_smooth"
[26] "stat_spoke"          "stat_sum"            "stat_summary"        "stat_summary_2d"      "stat_summary_bin"
[31] "stat_summary_hex"    "stat_summary2d"      "stat_unique"         "stat_ydensity"        "StatBin"
[36] "StatBin2d"           "StatBindot"          "StatBinhex"          "StatBoxplot"          "StatContour"
[41] "StatContourFilled"   "StatCount"           "StatDensity"         "StatDensity2d"        "state.abb"
[46] "state.area"          "state.center"        "state.division"      "state.name"           "state.region"
[51] "state.x77"           "StatEcdf"            "StatEllipse"         "StatFunction"         "StatIdentity"
[56] "StatQq"              "StatQqLine"          "StatQuantile"        "StatSf"               "StatSfCoordinates"
[61] "StatSmooth"          "StatSum"             "StatSummary"         "StatSummary2d"        "StatSummaryBin"
[66] "StatSummaryHex"      "StatUnique"          "StatYdensity"
```

Menyimpan Grafik

ggsave()

- saves last plot displayed
- requires file name to be supplied
- uses file name extension to determine file type:
.ps   .eps   .tex   .pdf   .jpg   .tiff   .png   .bmp   .svg   .wmf (windows only)
- uses size of current graphics device for default size

```
w <- read.csv(file="WDS2012.csv", head=TRUE, sep=",")
ggplot(data=w, aes(x=le, y=tfr, color=area)) + geom_point()

ggsave(file="le_tfr1.jpg")
ggsave(file="le_tfr2.jpg", scale=2)
ggsave(file="le_tfr3.jpg", width=5, height=5, unit="in")

ggsave(file="le_tfr4.png")
ggsave(file="le_tfr5.pdf")
```

Akhir materi 4b...