# Data visualization with ggplot2 Cheat Sheet

Gustavo Santos Paiva Laender Moura

2025-03-01



library(ggplot2)

# **Overview**

• Run ?ggplot2 for author information. ggplot2 is a system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

- For package documentation, run help(package = "ggplot2")
- Need help getting started? Try the R Graphics Cookbook
- Tidyverse::ggplot2
- Github
- Cheat sheet
- Cheat sheet pdf
- Run ??ggplot2 for broader search
- For an html helper page on your browser, run,

```
help.start()
```

If the browser launched by '/usr/bin/open' is already running, it is \*not\* restarted, and you must switch to its window.

Otherwise, be patient ...

```
browseVignettes("ggplot2")
```

Index

```
?ggplot2
??ggplot2
```

#### **Datasets**

The datasets used in the examples are: fruit (80 fruit names), words (980 words), and sentences (720 sentences). Run ?fruit for documentation.

#### **Basics**

ggplot2 is based on the grammar of graphics, the idea that you can build every graph from the same components: a data set, a coordinate system, and geoms—visual marks that represent data points.

### library(ggplot2)

To display values, map variables in the data to visual properties of the geom (aesthetics) like size, color, and x and y locations.

Complete the template below to build a graph.

Data, a Geom Function, and Aes Mappings are required. Stat, Position, and the Coordinate, Facet, Scale, and Theme functions are not required and will supply sensible defaults.

- ggplot(data = mpg, aes(x = cty, y = hwy)): Begins a plot that you finish by adding layers to. Add one geom function per layer.
- last\_plot(): Returns the last plot.
- ggsave("plot.png", width = 5, height = 5): Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

#### Aes

Common aesthetic values.

- color and fill: String ("red", "#RRGGBB").
- linetype: Integer or string (0 = "blank", 1 = "solid", 2 = "dashed", 3 = "dotted", 4 = "dotdash", 5 = "longdash", 6 = "twodash").
- size: Integer (in mm for size of points and text).

- linewidth: Integer (in mm for widths of lines).
- shape: Integer/shape name or a single character ("a").

```
- shape integer/name pairs: 0 = "square open", 1 = "circle open", 2 = "triangle open", 3 = "plus", 4 = "cross", 5 = "diamond open", 6 = "triangle down open", 7 = "square cross", 8 = "asterisk", 9 = "diamond plus", 10 = "circle plus", 11 = "star", 12 = "square plus", 13 = "circle cross", 14 = "square triangle", 15 = "square", 16 = "circle", 17 = "triangle", 18 = "diamond", 19 = "circle small", 20 = "bullet", 21 = "circle filled", 22 = "square filled", 23 = "diamond filled", 24 = "triangle filled", 25 = "triangle down filled"
```

#### Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

## **Graphical Primitives**

```
a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))</pre>
```

- a + geom\_blank() and a + expand\_limits(): Ensure limits include values across all plots.
- b + geom\_curve(aes(yend = lat + 1, xend = long + 1), curvature = 1): Draw a curved line from (x, y) to (xend, yend). aes() arguments: x, xend, y, yend, alpha, angle, color, curvature, linetype, size.
- a + geom\_path(lineend = "butt", linejoin = "round", linemitre = 1): Connect observations in the order they appear. aes() arguments: x, y, alpha, color, group, linetype, size.
- a + geom\_polygon(aes(alpha = 50)): Connect points into polygons. aes() arguments: x, y, alpha, color, fill, group, subgroup, linetype, size.
- b + geom\_rect(aes(xmin = long, ymin = lat, xmax = long + 1, ymax = lat + 1)): Draw a rectangle by connecting four corners (xmin, xmax, ymin, ymax). aes() arguments: xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size.

• a + geom\_ribbon(aes(ymin = unemploy - 900, ymax = unemploy + 900): For each x, plot an interval from ymin to ymax. aes() arguments: x, ymax, ymin, alpha, color, fill, group, linetype, size.

#### **Line Segments**

Common aesthetics: x, y, alpha, color, linetype, size, linewidth.

- b + geom\_abline(aes(intercept = 0, slope = 1)): Draw a diagonal reference line with a given slope and intercept.
- b + geom\_hline(aes(yintercept = lat)): Draw a horizontal reference line with a given yintercept.
- b + geom\_vline(aes(xintercept = long)): Draw a vertical reference line with a given xintercept.
- b + geom\_segment(aes(yend = lat + 1, xend = long + 1)): Draw a straight line from (x, y) to (xend, yend).
- b + geom\_spoke(aes(angle = 1:1155, radius = 1)): Draw line segments using polar coordinates (angle and radius).

#### One Variable - Continuous

```
c <- ggplot(mpg, aes(hwy))
c2 <- ggplot(mpg)</pre>
```

- c + geom\_area(stat = "bin"): Draw an area plot. aes() arguments: x, y, alpha, color, fill, linetype, linewidth.
- c + geom\_density(kernel = "gaussian"): Compute and draw kernel density estimates. aes() arguments: x, y, alpha, color, fill, group, linetype, linewidth, weight.
- c + geom\_dotplot(): Draw a dot plot. aes() arguments: x, y, alpha, color, fill.
- c + geom\_freqpoly(): Draw a frequency polygon. aes() arguments: x, y, alpha, color, group, linetype, linewidth.
- c + geom\_histogram(binwidth = 5): Draw a histogram. aes() arguments: x, y, alpha, color, fill, linetype, linewidth, weight.
- c2 + geom\_qq(aes(sample = hwy)): Draw a quantile-quantile plot. aes() arguments: x, y, alpha, color, fill, linetype, size, weight.

#### One Variable - Discrete

```
d <- ggplot(mpg, aes(fl))</pre>
```

• d + geom\_bar(): Draw a bar chart. aes() arguments: x, alpha, color, fill, linetype, linewidth, weight.

#### Two Variables - Both Continuous

```
e <- ggplot(mpg, aes(cty, hwy))
```

- e + geom\_label(aes(label = cty), nudge\_x = 1, nudge\_y = 1): Add text with a rectangle background. aes() arguments: x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust.
- e + geom\_point(): Draw a scatter plot. aes() arguments: x, y, alpha, color, fill, shape, size, stroke.
- e + geom\_quantile(): Fit and draw quantile regression for the plot data. aes() arguments: x, y, alpha, color, group, linetype, linewidth, weight.
- e + geom\_rug(sides = "bl"): Draw a rug plot. aes() arguments: x, y, alpha, color, linetype, linewidth.
- e + geom\_smooth(method = lm): Plot smoothed conditional means. aes() arguments: x, y, alpha, color, fill, group, linetype, linewidth, weight.
- e + geom\_text(aes(label = cty), nudge\_x = 1, nudge\_y = 1): Add text to a plot. aes() arguments: x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust.

#### Two Variables - One Discrete, One Continuous

```
f <- ggplot(mpg, aes(class, hwy))
```

- f + geom\_col(): Draw a bar plot. aes() arguments: x, y, alpha, color, fill, group, linetype, linewidth.
- f + geom\_boxplot(): Draw a box plot. aes() arguments: x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, linewidth, weight.

- f + geom\_dotplot(binaxis ="y", stackdir = "center"): Draw a dot plot. aes() arguments: x, y, alpha, color, fill, group.
- f + geom\_violin(scale = "area"): Draw a violin plot. aes() arguments: x, y, alpha, color, fill, group, linetype, linewidth, weight.

#### Two Variables - Both Discrete

```
g <- ggplot(diamonds, aes(cut, color))
```

- g + geom\_count(): Plot a count of points in an area to address over plotting. aes() arguments: x, y, alpha, color, fill, shape, size, stroke.
- e + geom\_jitter(height = 2, width = 2): Jitter points in a plot. aes() arguments: x, y, alpha, color, fill, shape, size.

#### Two Variables - Continuous Bivariate Distribution

```
h <- ggplot(diamonds, aes(carat, price))</pre>
```

- h + geom\_bin2d(binwidth = c(0.25, 500)): Draw a heatmap of 2D rectangular bin counts. aes() arguments: x, y, alpha, color, fill, linetype, size, weight.
- h + geom\_density\_2d(): Plot contours from 2D kernel density estimation. aes() arguments: x, y, alpha, color, group, linetype, linewidth.
- h + geom\_hex(): Draw a heatmap of 2D hexagonal bin counts. aes() arguments: x, y, alpha, color, fill, linewidth.

#### Two Variables - Continuous Function

```
i <- ggplot(economics, aes(date, unemploy))</pre>
```

- i + geom\_area(): Draw an area plot. aes() arguments: x, y, alpha, color, fill, linetype, linewidth.
- i + geom\_line(): Connect data points, ordered by the x axis variable. aes() arguments: x, y, alpha, color, group, linetype, linewidth.
- i + geom\_step(direction = "hv": Draw a stairstep plot. aes() arguments: x, y, alpha, color, group, linetype, linewidth.

#### Two Variables - Visualizing Error

```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit - se, ymax = fit + se))</pre>
```

- j + geom\_crossbar(fatten = 2): Draw a crossbar. aes() arguments: x, y, ymax, ymin, alpha, color, fill, group, linetype, linewidth.
- j + geom\_errorbar(): Draw an errorbar. Also geom\_errorbarh(). aes() arguments: x, ymax, ymin, alpha, color, group, linetype, linewidth, width.
- j + geom\_linerange(): Draw a line range. aes() arguments: x, ymin, ymax, alpha, color, group, linetype, linewidth.
- j + geom\_pointrange(): Draw a point range. aes() arguments: x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, linewidth.

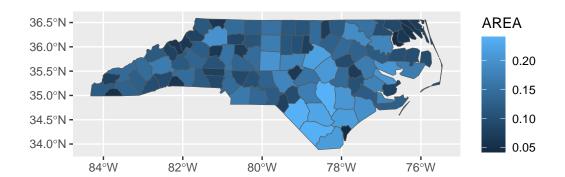
## Two Variables - Maps

Draw the appropriate geometric object depending on the simple features present in the data. aes() arguments: map\_id, alpha, color, fill, linetype, linewidth.

```
nc <- sf::st_read(system.file("shape/nc.shp", package = "sf"))

Reading layer `nc' from data source
   `/Users/gustavosplmoura/Library/R/arm64/4.4/library/sf/shape/nc.shp'
   using driver `ESRI Shapefile'
Simple feature collection with 100 features and 14 fields
Geometry type: MULTIPOLYGON
Dimension: XY
Bounding box: xmin: -84.32385 ymin: 33.88199 xmax: -75.45698 ymax: 36.58965
Geodetic CRS: NAD27</pre>

ggplot(nc) +
   geom_sf(aes(fill = AREA))
```



#### Three Variables

```
seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2))
1 <- ggplot(seals, aes(long, lat))</pre>
```

- 1 + geom\_contour(aes(z = z)): Draw 2D contour plot. aes() arguments: x, y, z, alpha, color, group, linetype, linewidth, weight.
- 1 + geom\_contour\_filled(aes(fill = z)): Draw 2D contour plot with the space between lines filled. aes() arguments: x, y, alpha, color, fill, group, linetype, linewidth, subgroup.
- 1 + geom\_raster(aes(fill = z), hjust = 0.5, vjust = 0.5, interpolate = FALSE): Draw a raster plot. aes() arguments: x, y, alpha, fill.
- 1 + geom\_tile(aes(fill = z)): Draw a tile plot. aes() arguments: x, y, alpha, color, fill, linetype, linewidth, width.

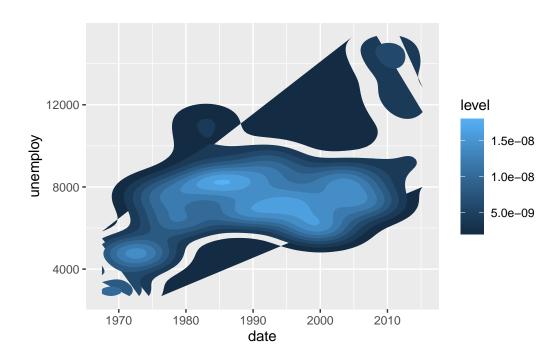
#### Stats

An alternative way to build a layer.

A stat builds new variables to plot (e.g., count, prop).

Visualize a stat by changing the default stat of a geom function, geom\_bar(stat = "count"), or by using a stat function, stat\_count(geom = "bar"), which calls a default geom to make a layer (equivalent to a geom function). Use after\_stat(name) syntax to map the stat variable name to an aesthetic.

#### i + stat\_density\_2d(aes(fill = after\_stat(level)), geom = "polygon")



In this example, "polygon" is the geom to use, stat\_density\_2d() is the stat function, aes() contains the geom mappings, and level is the variable created by stat.

- c + stat\_bin(binwidth = 1, boundary = 10): x, y | count, ncount, density, ndensity
- c + stat\_count(width = 1): x, y | count, density
- c + stat\_density(adjust = 1, kernel = "gaussian"): x, y | count, density, scaled
- e + stat\_bin\_2d(bins = 30, drop = T): x, y, fill | count, density
- e + stat\_bin\_hex(bins =30): x, y, fill | count, density
- e + stat\_density\_2d(contour = TRUE, n = 100): x, y, color, linewidth | level
- e + stat\_ellipse(level = 0.95, segments = 51, type = "t")

```
• 1 + stat_contour(aes(z = z)): x, y, z, order | level
```

- 1 + stat\_summary\_hex(aes(z = z), bins = 30, fun = max): x, y, z, fill | value
- 1 + stat\_summary\_2d(aes(z = z), bins = 30, fun = mean):  $x, y, z, fill \mid value$
- f + stat\_boxplot(coef = 1.5): x, y | lower, middle, upper, width, ymin, ymax
- f + stat\_ydensity(kernel = "gaussian", scale = "area"): x, y | density, scaled, count, n, violinwidth, width
- e + stat\_ecdf(n = 40): x, y | x, y
- e + stat\_quantile(quantiles = c(0.1, 0.9), formula = y ~ log(x), method = "rq"): x, y | quantile
- e + stat\_smooth(method = "lm", formula = y ~ x, se = T, level = 0.95): x,
   y | se, x, y, ymin, ymax
- ggplot() + xlim(-5, 5) + stat\_function(fun = dnorm, n = 20, geom =
   "point"): x | x, y
- ggplot() + stat\_qq(aes(sample = 1:100)): x, y, sample | sample, theoretical
- e + stat\_sum(): x, y, size | n, prop
- e + stat\_summary(fun.data = "mean\_cl\_boot")
- h + stat\_summary\_bin(fun = "mean", geom = "bar")
- e + stat\_identity()
- e + stat\_unique()

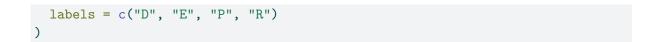
#### Scales

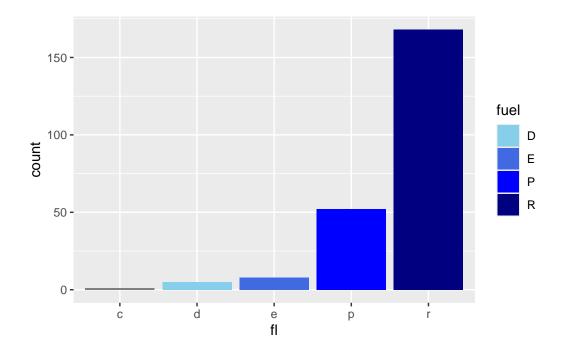
Override defaults with **scales** package.

**Scales** map data values to the visual values of an aesthetic. To change a mapping, add a new scale.

```
n <- d + geom_bar(aes(fill = fl))

n + scale_fill_manual(
  values = c("skyblue", "royalblue", "blue", "navy"),
  limits = c("d", "e", "p", "r"),
  breaks =c("d", "e", "p", "r"),
  name = "fuel",</pre>
```





In this example, scale\_ specifies a scale function, fill is the aesthetic to adjust, and manual is the prepackaged scale to use.

values contains scale-specific arguments, limits specifies the range of values to include in mappings, breaks specifies the breaks to use in legend/axis, and name and labels specify the title and labels to use in the legend/axis.

## **General Purpose Scales**

Use with most aesthetics.

- scale\_\*\_continuous(): Map continuous values to visual ones.
- scale\_\*\_discrete(): Map discrete values to visual ones.
- scale\_\*\_binned(): Map continuous values to discrete bins.
- scale\_\*\_identity(): Use data values as visual ones.
- scale\_\*\_manual(values = c()): Map discrete values to manually chosen visual ones.

- scale\_\*\_date(date\_labels = "%m/%d", date\_breaks = "2 weeks"): Treat data values as dates.
- scale\_\*\_datetime(): Treat data values as date times. Same as scale\_\*\_date(). See ?strptime for label formats.

#### X & Y Location Scales

Use with x or y aesthetics (x shown here).

- scale\_x\_log10(): Plot x on log10 scale.
- scale\_x\_reverse(): Reverse the direction of the x axis.
- scale\_x\_sqrt(): Plot x on square root scale.

## Color and Fill Scales (Discrete)

- n + scale\_fill\_brewer(palette = "Blues"): Use color scales from ColorBrewer. For palette choices RColorBrewer::display.brewer.all().
- n + scale\_fill\_grey(start = 0.2, end = 0.8, na.value = "red"): Use a grey gradient color scale.

## Color and Fill Scales (Continuous)

```
o <- c + geom_dotplot(aes(fill = ..x..))
```

- o + scale\_fill\_distiller(palette = "Blues"): Interpolate a palette into a continuous scale.
- o + scale\_fill\_gradient(low = "red", high = "yellow"): Create a two color gradient.
- o + scale\_fill\_gradient2(low = "red", high = "blue", mid = "white", midpoint = 25): Create a diverging color gradient.
- o + scale\_fill\_gradientn(colors = topo.colors(6)): Create a n-color gradient.
  Also rainbow(), heat.colors(), terrain.colors(), cm.colors(), RColorBrewer::brewer.pal().

#### Shape and Size Scales

```
p <- e + geom_point(aes(shape = fl, size = cyl))</pre>
```

- p + scale\_shape() + scale\_size(): Map discrete values to shape and size aesthetics.
- p + scale\_shape\_manual(values = c(3:7)): Map discrete values to specified shape values.
- p + scale\_radius(range = c(1,6)): Map values to a shape's radius.
- p + scale\_size\_area(max\_size = 6): Like scale\_size() but maps zero values to zero size.

Shapes used here are the same as the ones listed in the Aes section.

## **Coordinate Systems**

```
u <- d + geom_bar()
```

- u + coord\_cartesian(xlim = c(0, 5)): xlim, ylim. The default Cartesian coordinate system.
- u + coord\_fixed(ratio = 1/2): ratio, xlim, ylim. Cartesian coordinates with fixed aspect ration between x and y units.
- ggplot(mpg, aes(y = fl)) + geom\_bar(): Flip Cartesian coordinates by switching x and y aesthetic mappings.
- u + coord\_polar(theta = "x", direction = 1): theta, start, direction. Polar coordinates.
- u + coord\_trans(y = "sqrt"): x, y, xlim, ylim. Transformed Cartesian coordinates. Set xtrans and ytrans to the name of a window function.
- + coord\_sf(): xlim, ylim, crs. Ensures all layers use a common Coordinate Reference System.

## **Position Adjustments**

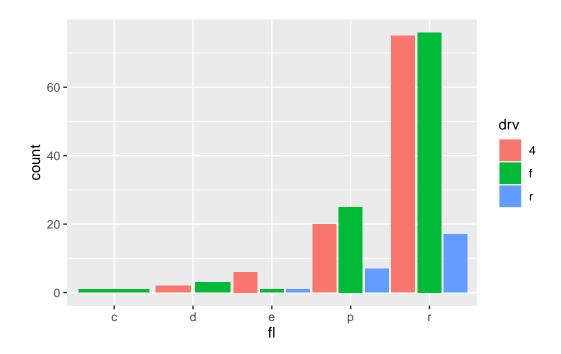
Position adjustments determine how to arrange geoms that would otherwise occupy the same space.

```
s <- ggplot(mpg, aes(fl, fill = drv))</pre>
```

- s + geom\_bar(position = "dodge"): Arrange elements side by side.
- s + geom\_bar(position = "fill"): Stack elements on top of one another, normalize height.
- e + geom\_point(position = "jitter"): Add random noise to X and Y position of each element to avoid over plotting.
- e + geom\_label(position = "nudge"): Nudge labels away from points.
- s + geom\_bar(position = "stack"): Stack elements on top of one another.

Each position adjustment can be recast as a function with manual width and height arguments:

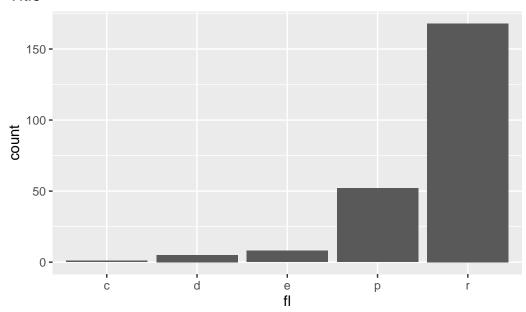
## s + geom\_bar(position = position\_dodge(width = 1))



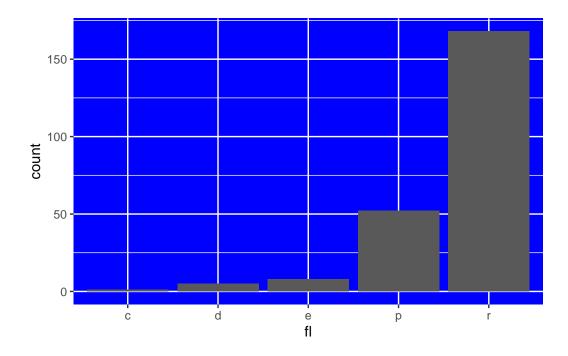
## **Themes**

- u + theme\_bw(): White background with grid lines.
- u + theme\_gray(): Grey background with white grid lines (default theme).
- u + theme\_dark(): Dark grey background and grid lines for contrast.
- u + theme\_classic(): No grid lines.
- u + theme\_light(): Light grey axes and grid lines.
- u + theme\_linedraw(): Uses only black lines.
- u + theme\_minimal(): Minimal theme.
- u + theme\_void(): Empty theme.
- u + theme(): Customize aspects of the theme such as axis, legend, panel, and facet properties.

## Title



```
u + theme(panel.background = element_rect(fill = "blue"))
```



## **Faceting**

Facets divide a plot into subplots based on the values of one or more discrete variables.

```
t <- ggplot(mpg, aes(cty, hwy)) + geom_point()
```

- t + facet\_grid(. ~ fl): Facet into a column based on fl.
- t + facet\_grid(year ~ .): Facet into rows based on year.
- t + facet\_grid(year ~ fl): Facet into both rows and columns.
- t + facet\_wrap(~ fl): Wrap facets into a rectangular layout.
- t + facet\_grid(drv ~ fl, scales = "free"): Set scales to let axis limits vary across facets. Also "free\_x" for x axis limits adjust to individual facets and "free\_y" for y axis limits adjust to individual facets.

Set labeller to adjust facet label:

- t + facet\_grid(. ~ fl, labeller = label\_both): Labels each facet as "fl: c", "fl: d", etc.
- t + facet\_grid(fl ~ ., labeller = label\_bquote(alpha ^ .(fl))): Labels each facet as "c", "d", etc.

## **Labels and Legends**

Use labs() to label elements of your plot.

```
t + labs(x = "New x axis label",
  y = "New y axis label",
  title ="Add a title above the plot",
  subtitle = "Add a subtitle below title",
  caption = "Add a caption below plot",
  alt = "Add alt text to the plot",
  <Aes> = "New <Aes> legend title")
```

- t + annotate(geom = "text", x = 8, y = 9, label = "A"): Places a geom with manually selected aesthetics.
- p + guides(x = guide\_axis(n.dodge = 2)): Avoid crowded or overlapping labels with guide\_axis(n.dodge or angle).
- n + guides(fill = "none"): Set legend type for each aesthetic: colorbar, legend, or none (no legend).
- n + theme(legend.position = "bottom"): Place legend at "bottom", "top", "left", or "right".
- n + scale\_fill\_discrete(name = "Title", labels = c("A", "B", "C", "D", "E")): Set legend title and labels with a scale function.

## Zooming

- t + coord\_cartesian(xlim = c(0, 100), ylim = c(10, 20)): Zoom without clipping (preferred).
- $t + x \lim(0, 100) + y \lim(10, 20)$  or  $t + s cale_x continuous(limits = c(0, 100))$ 100)) + scale\_y\_continuous(limits = c(0, 100)): Zoom with clipping (removes unseen data points).

CC BY SA Posit Software, PBC • info@posit.co • posit.co

Learn more at ggplot2.tidyverse.org.

Updated: 2025-03.

packageV	dersion(	("ggn]	ot.2"
packagev	CT DIOIL	\	1002 )

[1] '3.5.1'