

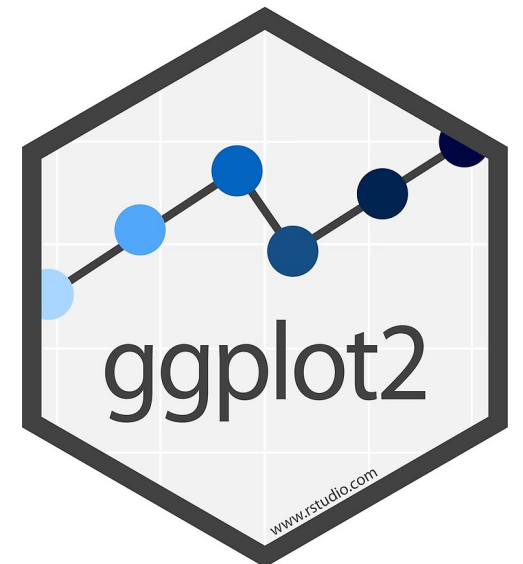
ggplot2: An extensible platform for publication-quality graphics

Claus O. Wilke

The University of Texas at Austin

 @clauswilke

 clauswilke





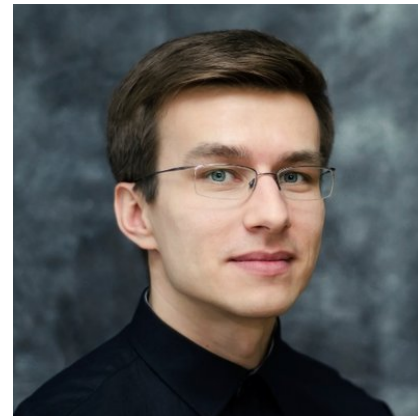
Hadley Wickham



Thomas Lin Pederson



Edzer Pebesma



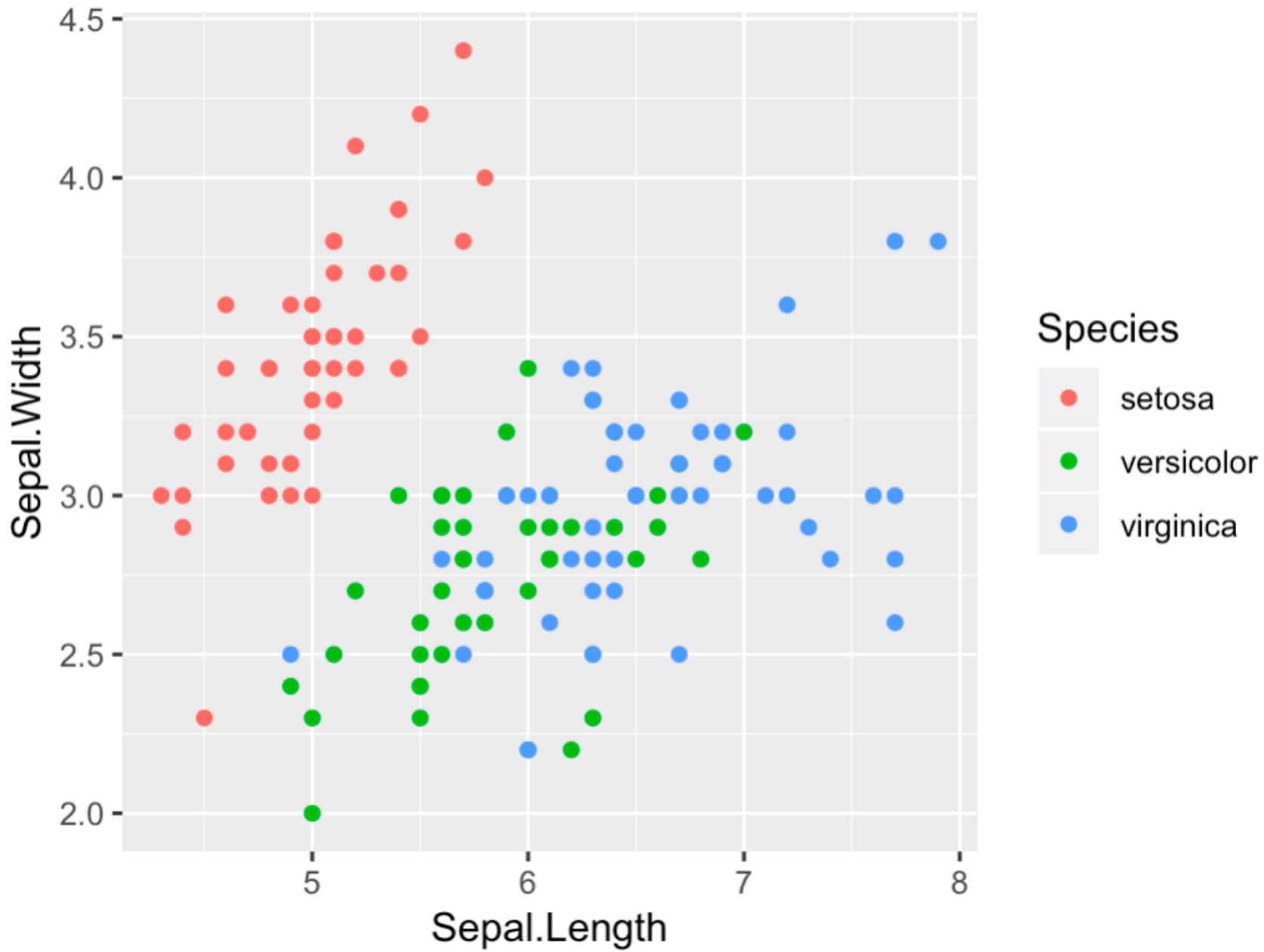
Kamil Slowikowski

2008

Practical tools for exploring data and models

Hadley Alexander Wickham
Iowa State University

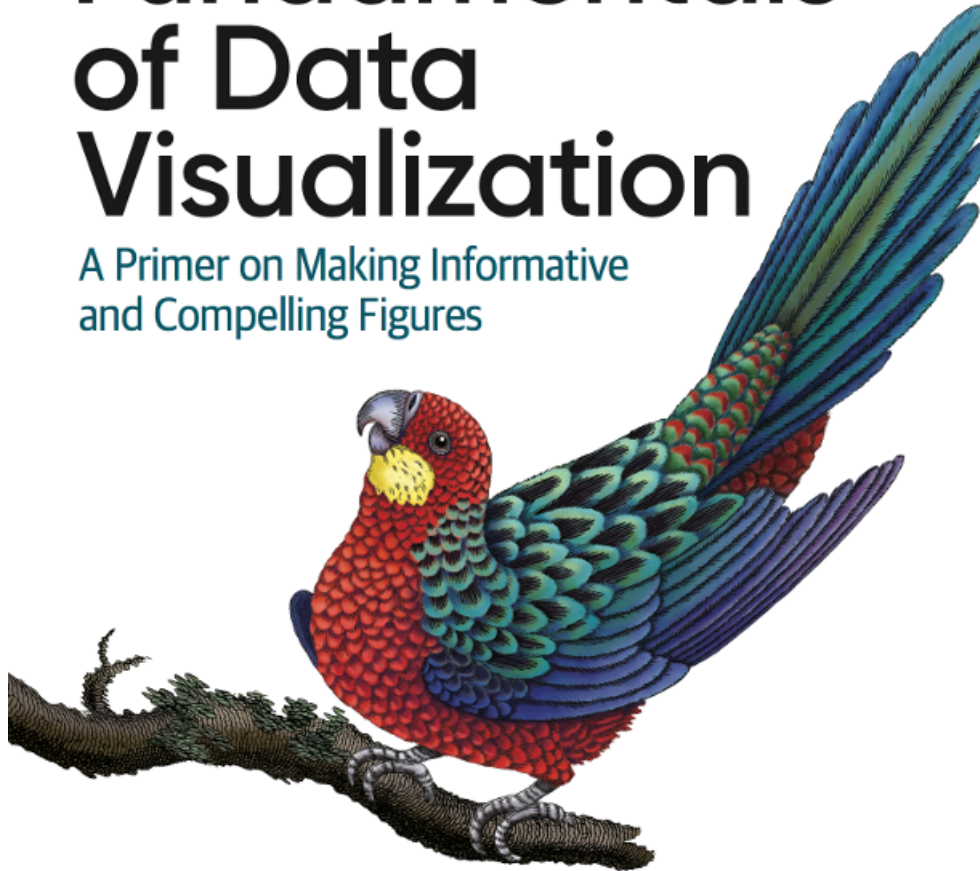




O'REILLY®

Fundamentals of Data Visualization

A Primer on Making Informative and Compelling Figures



Claus O. Wilke

serialmentor.com/dataviz

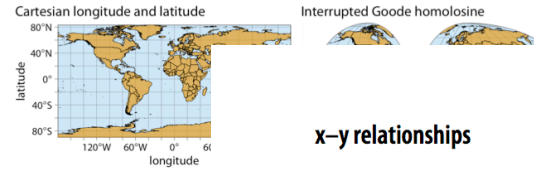
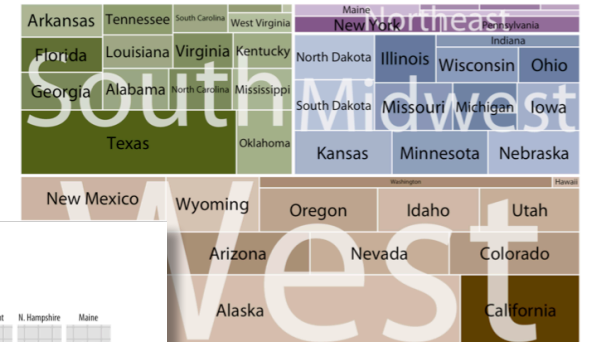


Figure 3-11. Map of the world, showing the Cartesian coordinate system. The angles relative to their true v sine projection perfectly represent masses into separate pieces, most jectio c distortic

x-y relationships



Scatterplots (Chapter 12) represent the arc show one quantitative variable relative to an axes, we can map one onto the dot size, cre bubble chart. For paired data, where the va ured in the same units, it is generally help "Paired Data" on page 127). Paired data can points connected by straight lines.



States in the US visualized as a treemap. Each rectangle represents one state. The coloring represents the region.

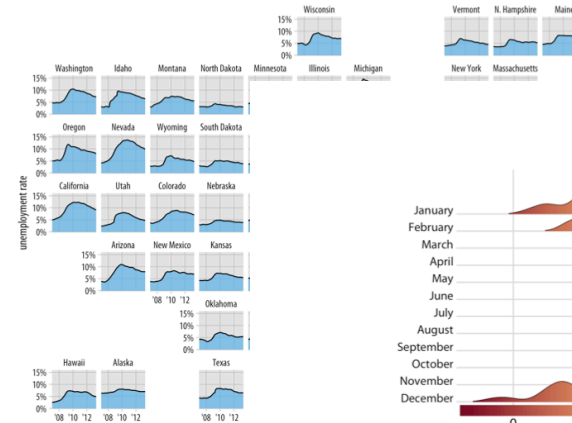


Figure 15-17. Unemployment rate lead by state. Each panel shows the unemplo Columbia (DC), from January 2007 thr of 2008, 2010, and 2012. States that are in the unemployment rate. Data source:

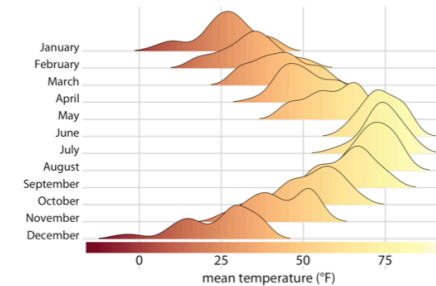


Figure 20-12. Temperatures in Lincoln, NE, in 2016. This figure is a variation of Figure 9-9. Temperature is now shown both by location along the x axis and by color, and a color bar along the x axis visualizes the scale that converts temperatures into colors. Data source: Weather Underground.

What do we need?

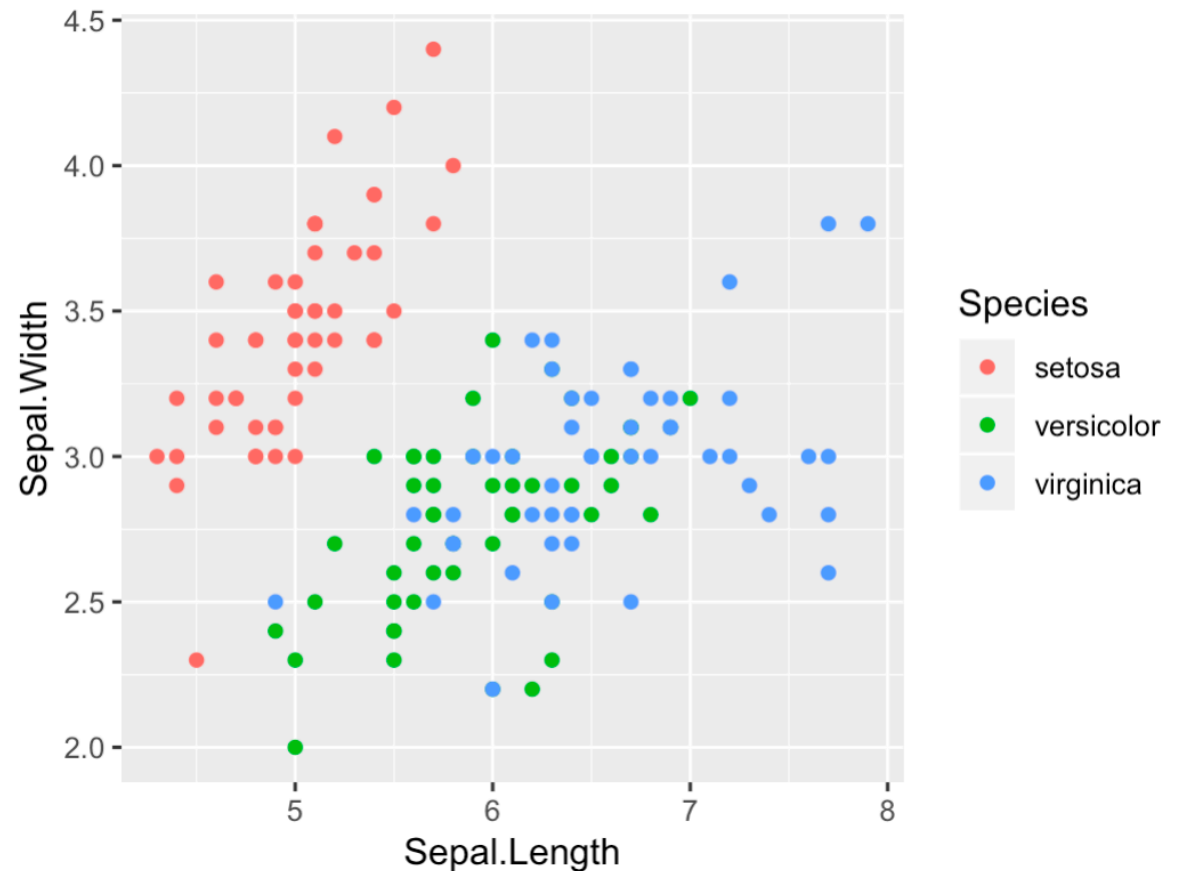
- Powerful styling options
- Broad selection of plot types
- Sophisticated text annotations
- Plot composition

What do we need?

- Powerful styling options
- Broad selection of plot types
- Sophisticated text annotations
- Plot composition

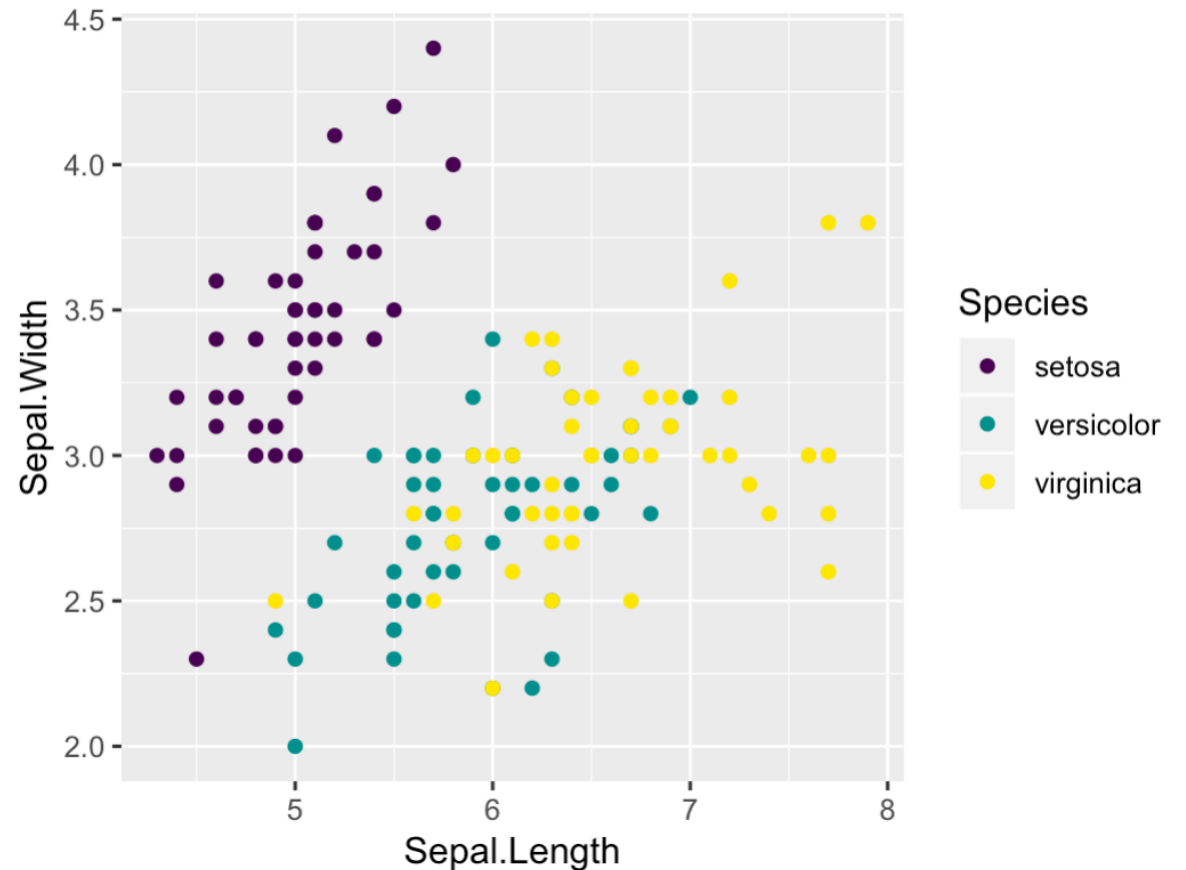
ggplot2: A grammar of graphics in R

```
ggplot(iris) +  
  aes(x = Sepal.Length, y = Sepal.Width, color = Species) +  
  geom_point()
```



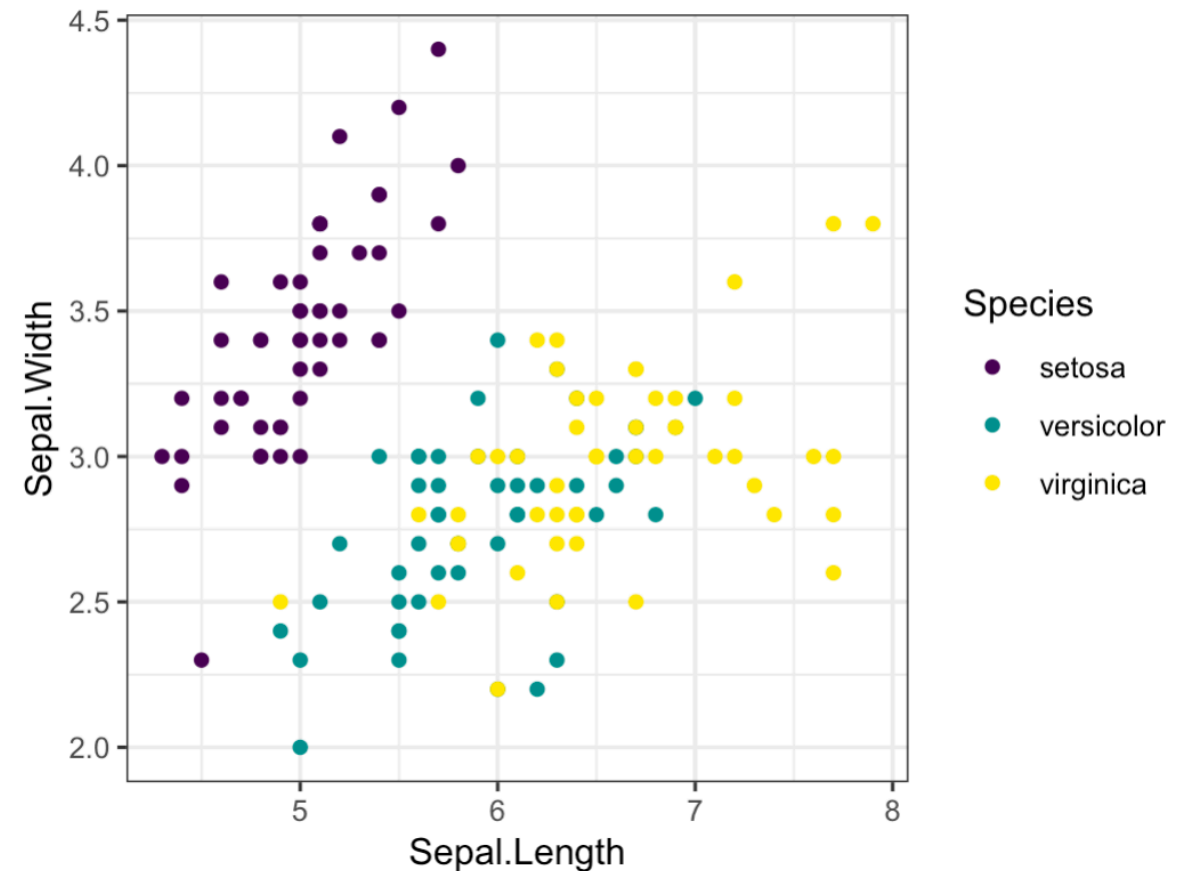
ggplot2: A grammar of graphics in R

```
ggplot(iris) +  
  aes(x = Sepal.Length, y = Sepal.Width, color = Species) +  
  geom_point() +  
  scale_color_viridis_d()
```



ggplot2: A grammar of graphics in R

```
ggplot(iris) +  
  aes(x = Sepal.Length, y = Sepal.Width, color = Species) +  
  geom_point() +  
  scale_color_viridis_d() +  
  theme_bw()
```



What do we need?

- Powerful styling options
- Broad selection of plot types
- Sophisticated text annotations
- Plot composition

52 registered extensions available to explore

Sort

Github stars

Text Filter

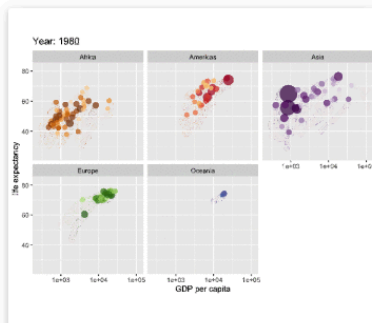
search name, author, de

Author Filter

Tag Filter

CRAN Only

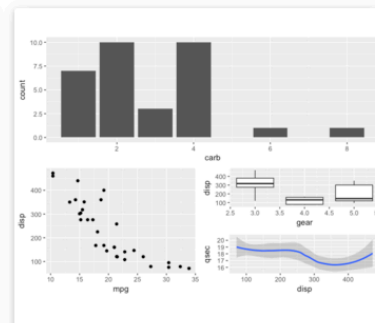
Showing 52 of 52

gganimate [Star 1185](#)

A Grammar of Animated Graphics.

author: [thomasp85](#)tags: [visualization](#), [general](#)

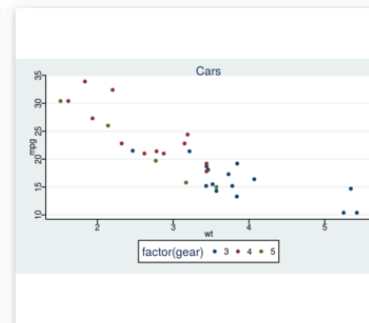
js libraries:

patchwork [Star 1017](#)

Easy composition of ggplot plots using arithmetic operators

author: [thomasp85](#)tags: [visualization](#), [composition](#)

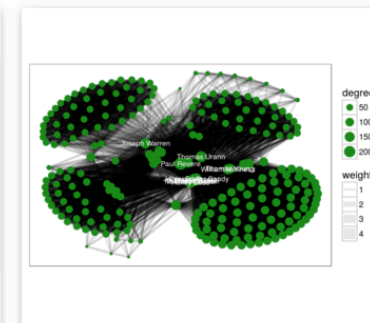
js libraries:

ggthemes [Star 967](#)

Some extra geoms, scales, and themes for ggplot.

author: [jmold](#)tags: [visualization](#), [general](#)

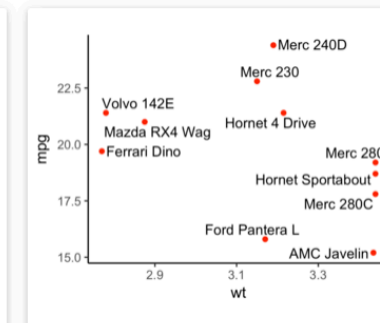
js libraries:

ggraph [Star 618](#)

ggraph is tailored at plotting graph-like data structures (graphs, networks, trees, hierarchies...).

author: [thomasp85](#)tags: [visualization](#), [general](#)

js libraries:

ggrepel [Star 600](#)

Repel overlapping text labels away from each other.

author: [slowkow](#)tags: [visualization](#), [general](#)

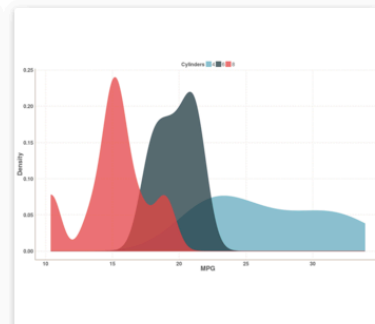
js libraries:

esquisse [Star 598](#)

Explore and Visualize Your Data Interactively with ggp1ot2

author: [dreamrs](#)tags: [visualization](#), [interface](#)

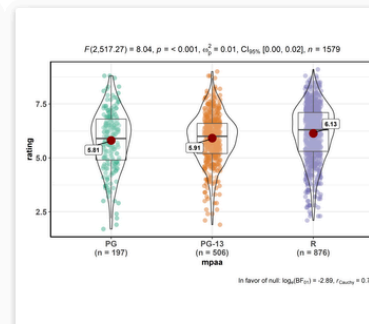
js libraries:

ggthemr [Star 512](#)

Themes for ggplot

author: [cttobin](#)tags: [visualization](#), [general](#)

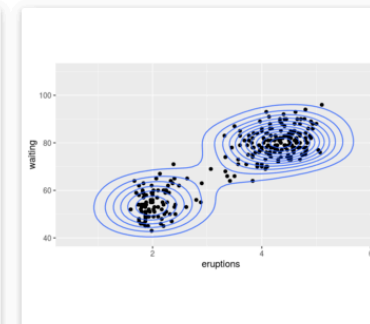
js libraries:

ggstatsplot [Star 448](#)

'ggstatsplot' provides a collection of functions to enhance 'ggplot2' plots with results from statistical tests.

author: [IndrajeetPatil](#)tags: [visualization](#), [statistics](#)

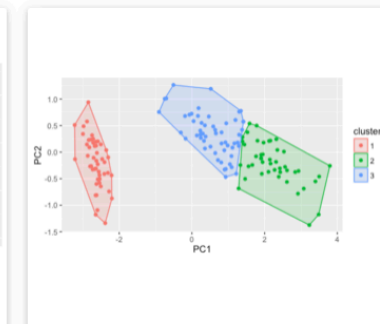
js libraries:

ggalt [Star 420](#)

A compendium of 'geoms', 'coords' and 'stats' for 'ggplot2'.

author: [hrbrmstr](#)tags: [visualization](#), [general](#)

js libraries:

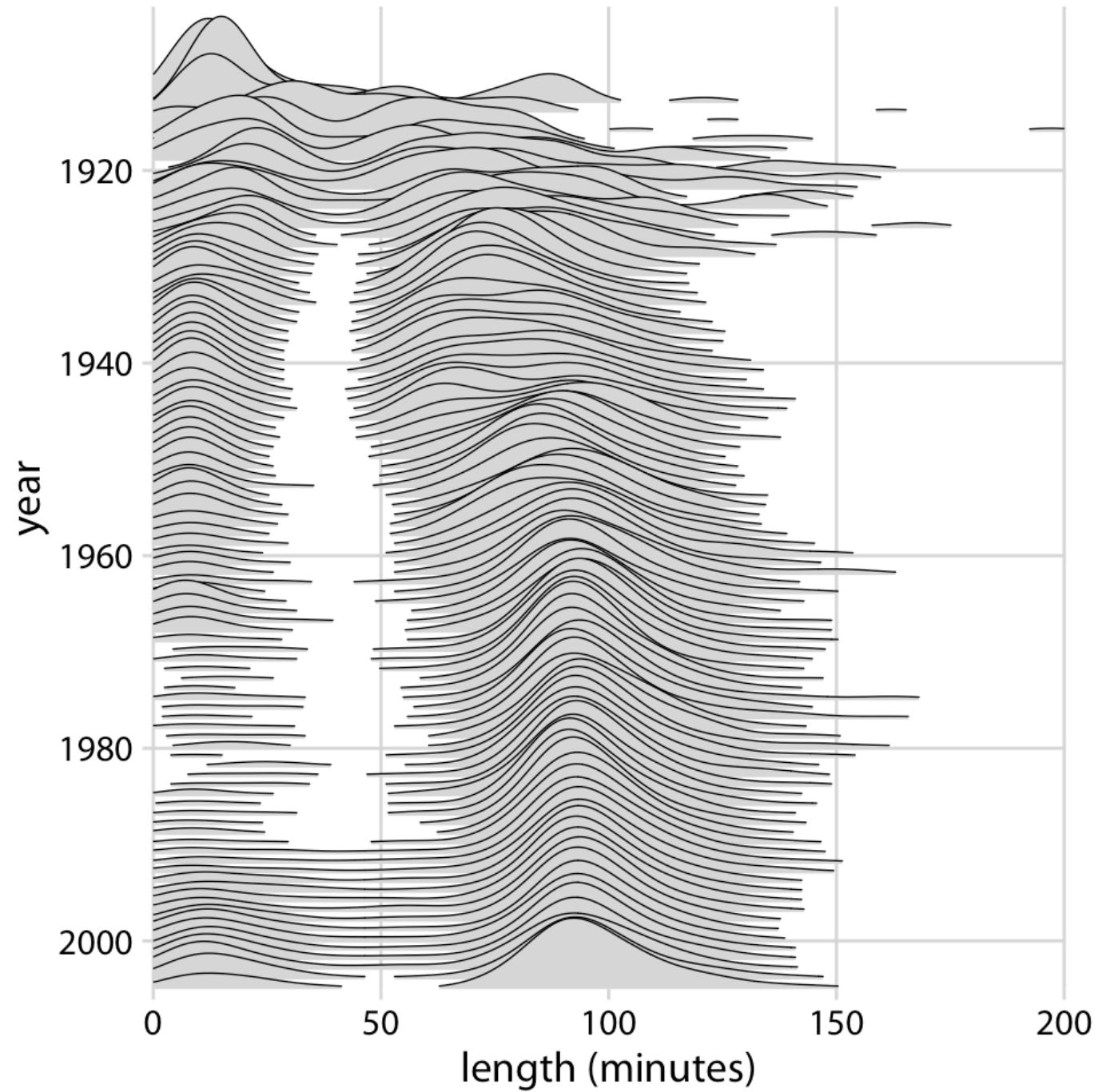
ggfortify [Star 394](#)

The unified interface to ggplot2 many popular statistical package results.

author: [terrytangyuan](#)tags: [visualization](#), [general](#)

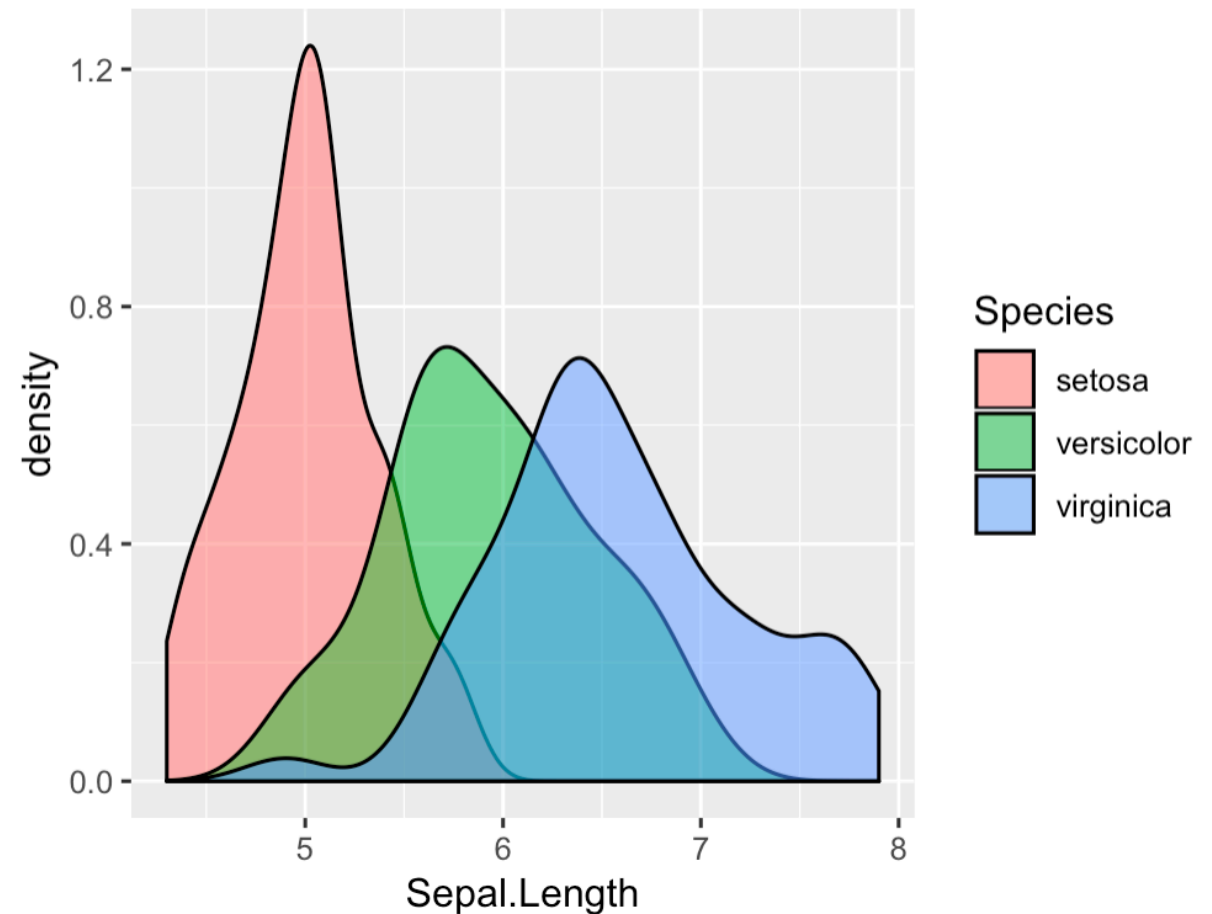
js libraries:

Example 1: Ridgeline plots



Ridgeline plots are a variation of density plots

```
ggplot(iris) +  
  aes(x = Sepal.Length, fill = Species) +  
  geom_density(alpha = 0.5)
```



Ridgeline plots are a variation of density plots

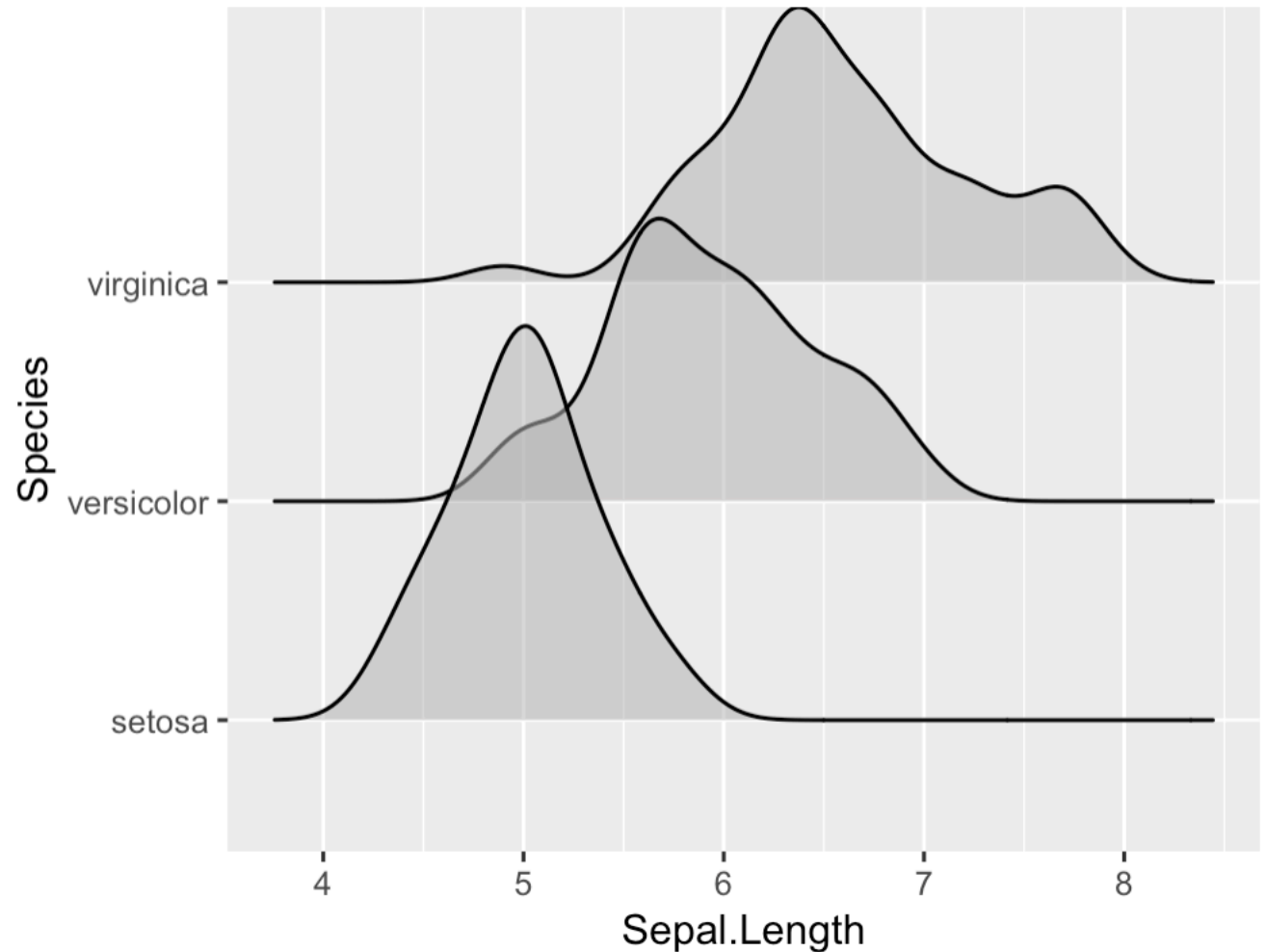
```
library(ggribes)
```

```
ggplot(iris) +  
  aes(x = Sepal.Length, y = Species) +  
  geom_density_ridges(alpha = 0.5)
```

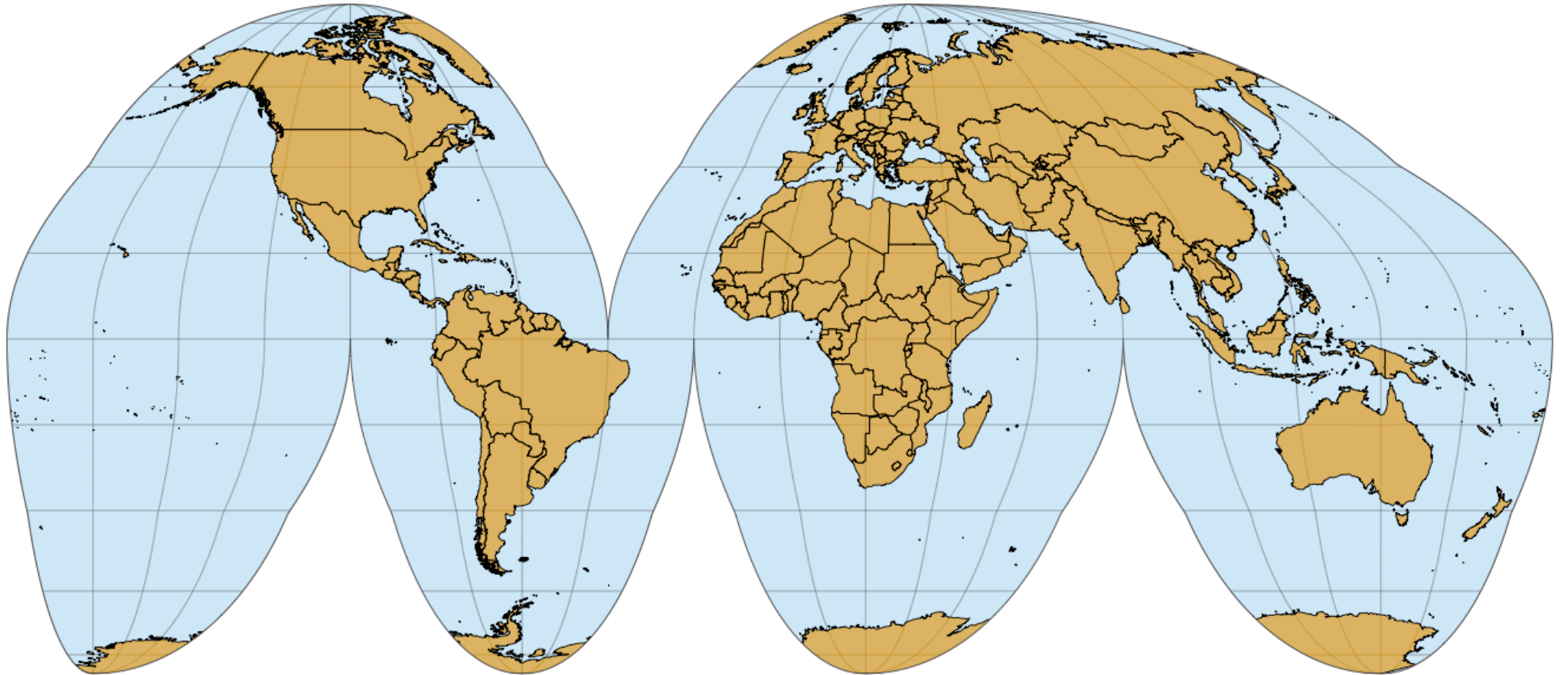
Ridgeline plots are a variation of density plots

```
library(ggribes)
```

```
ggplot(iris) +  
  aes(x = Sepal.Length,  
      geom_density_ridges(a
```

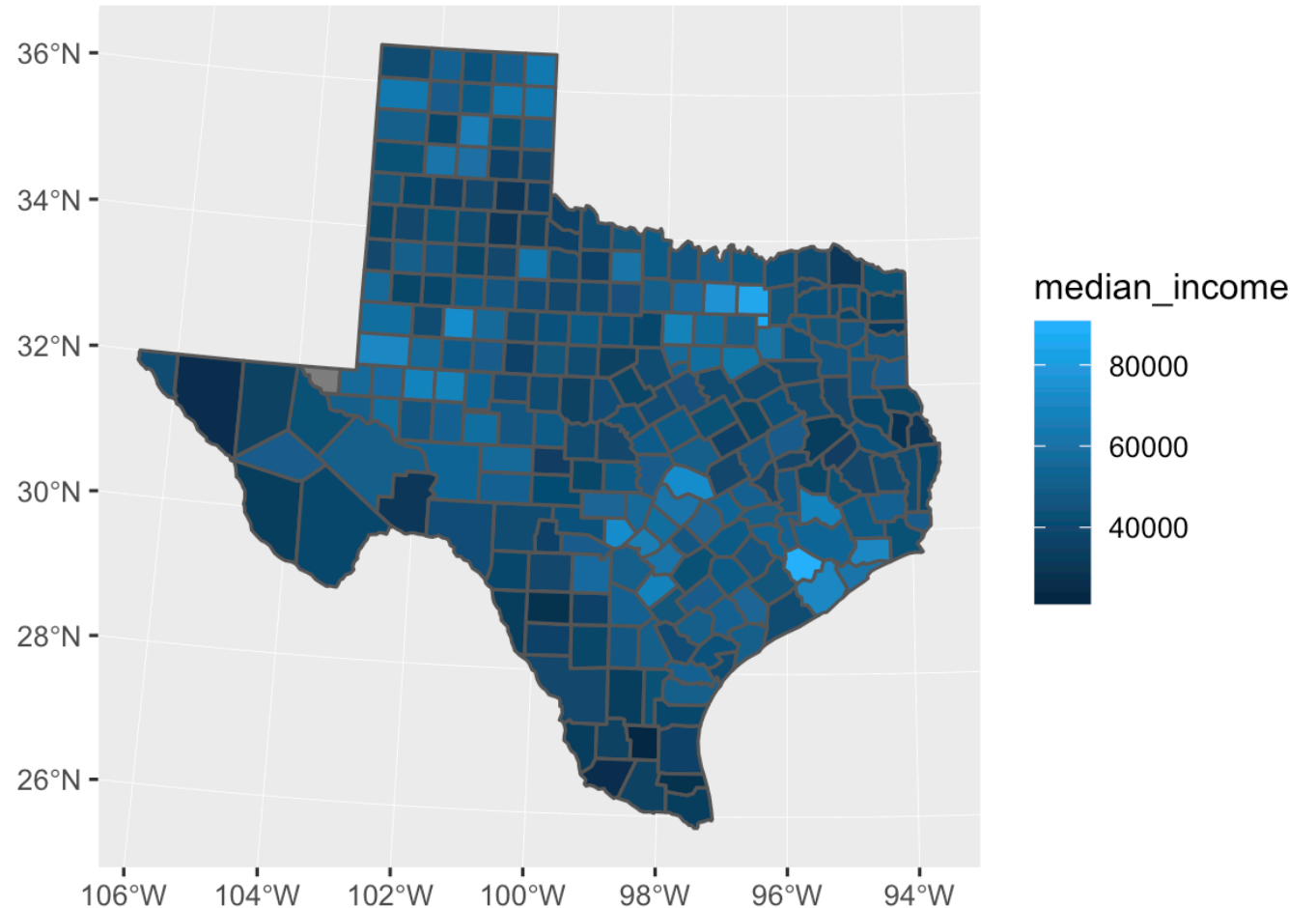


Example 2: Simple features



sf: Manipulating and plotting simple features

```
ggplot(TX_income, aes(fill = median_income)) +  
  geom_sf()
```



Edzer
Pebesma

TX_income[1:3,]

#> Simple feature collection with 3 features and 15 fields

#> geometry type: MULTIPOLYGON

#> dimension: XY

#> bbox: xmin: -579099 ymin: -958812.5 xmax: -168152.2 ymax: -94284.87

#> epsg (SRID): NA

#> proj4string: +proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=37.5 +lon_0=-96

+x_0=0 +y_0=0 +ellps=GRS80 +towgs84=0,0,0,0,0,0 +units=m +no_defs

#> STATEFP COUNTYFP COUNTYNS AFFGEOID GEOID NAME LSAD ALAND

#> 1 48 421 01383996 0500000US48421 48421 Sherman 06 2390651189

#> 2 48 493 01384032 0500000US48493 48493 Wilson 06 2081662847

#> 3 48 115 01383843 0500000US48115 48115 Dawson 06 2331781556

#> AWATER name median_income median_income_moe

#> 1 428754 Sherman County, Texas 51987 4386

#> 2 12111367 Wilson County, Texas 68100 3328

#> 3 4720730 Dawson County, Texas 41095 2591

#> population area popdens

#> 1 3066 2387929738 [m^2] 1.283957e-06 [1/m^2]

#> 2 45509 2086919506 [m^2] 2.180678e-05 [1/m^2]

#> 3 13542 2336898468 [m^2] 5.794860e-06 [1/m^2]

#> geometry

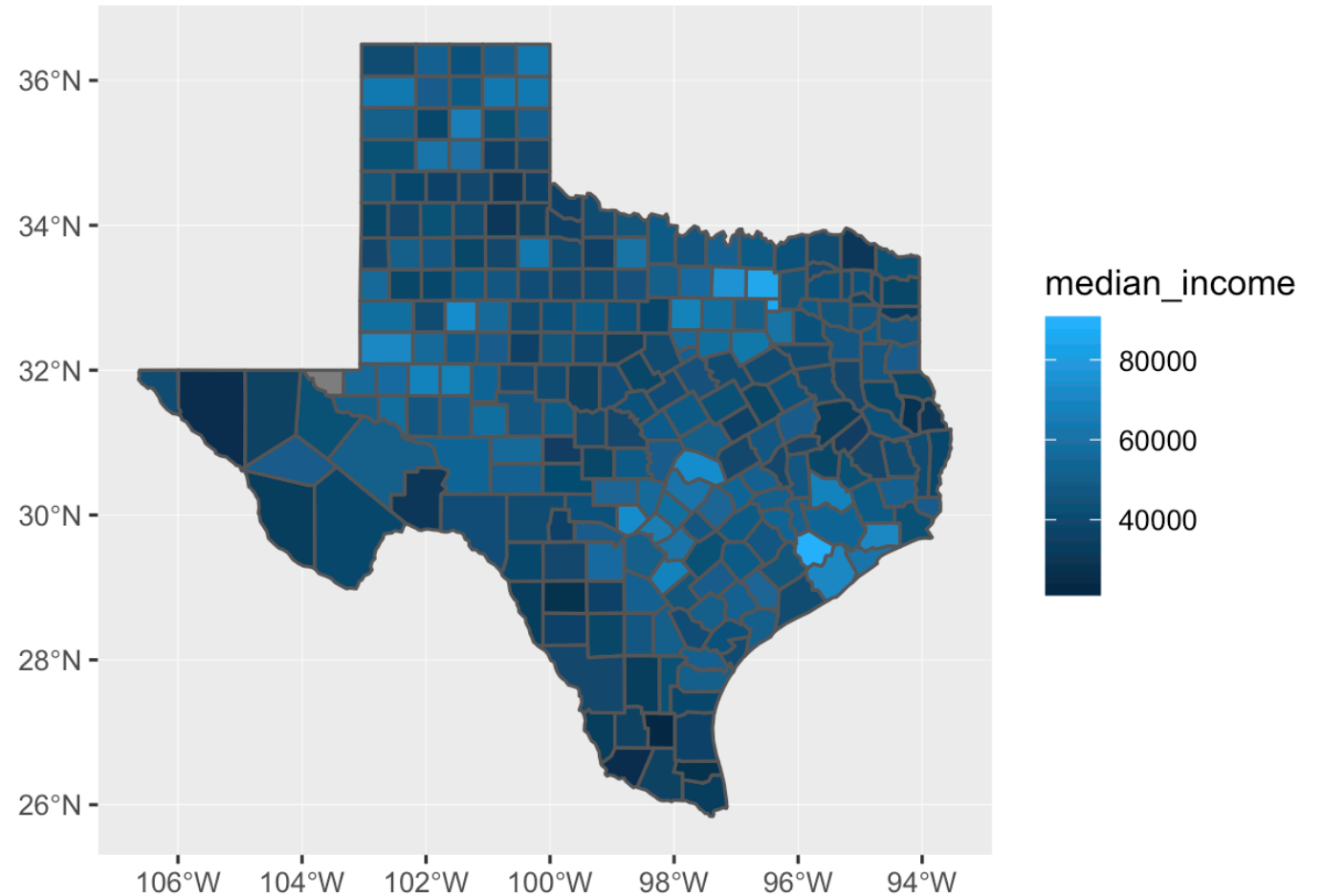
#> 1 MULTIPOLYGON (((-546533.1 -...

#> 2 MULTIPOLYGON (((-234487.5 -...

#> 3 MULTIPOLYGON (((-576468.3 -...

sf: Manipulating and plotting simple features

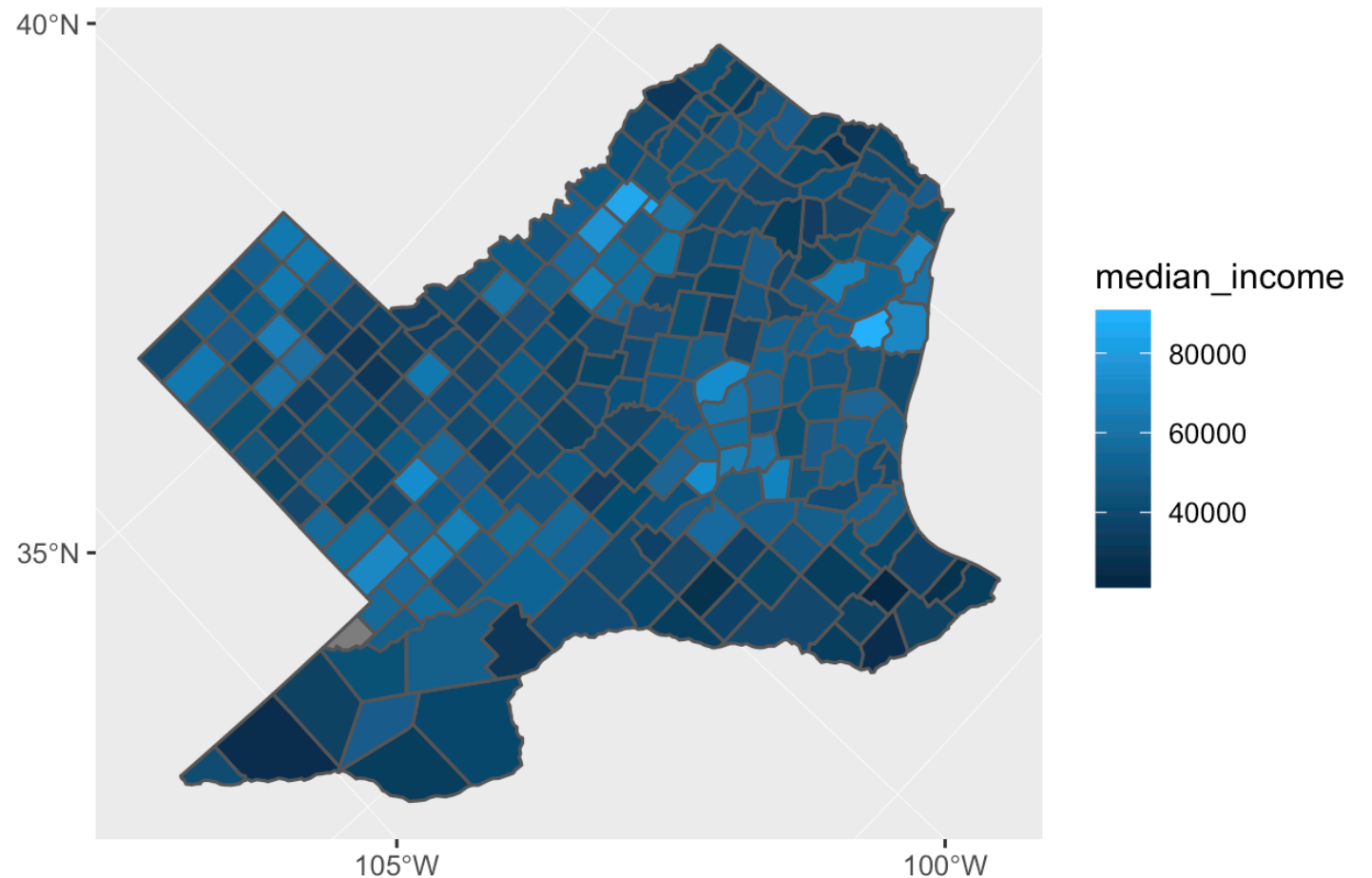
```
ggplot(TX_income, aes(fill = median_income)) +  
  geom_sf() +  
  coord_sf(crs = 4326) # Cartesian longitude and latitude
```



Edzer
Pebesma

sf: Manipulating and plotting simple features

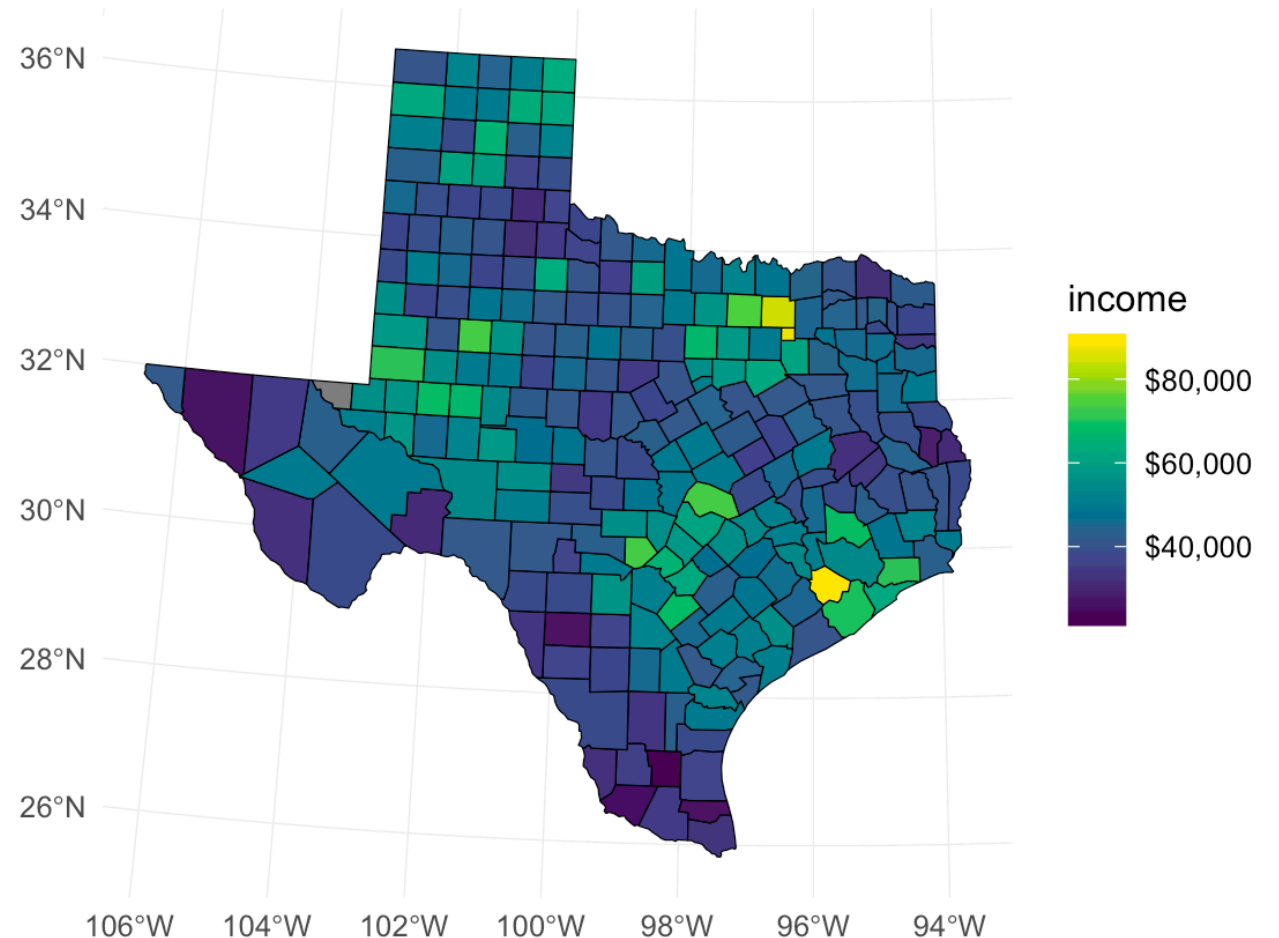
```
ggplot(TX_income, aes(fill = median_income)) +  
  geom_sf() +  
  coord_sf(crs = 3338) # NAD83, Alaska Albers
```



Edzer
Pebesma

sf: Manipulating and plotting simple features

```
# with formatting tweaks
ggplot(TX_income, aes(fill = median_income)) +
  geom_sf(color = "black", size = 0.2) +
  scale_fill_viridis_c(
    name = "income",
    labels = dollar
  ) +
  theme_minimal()
```



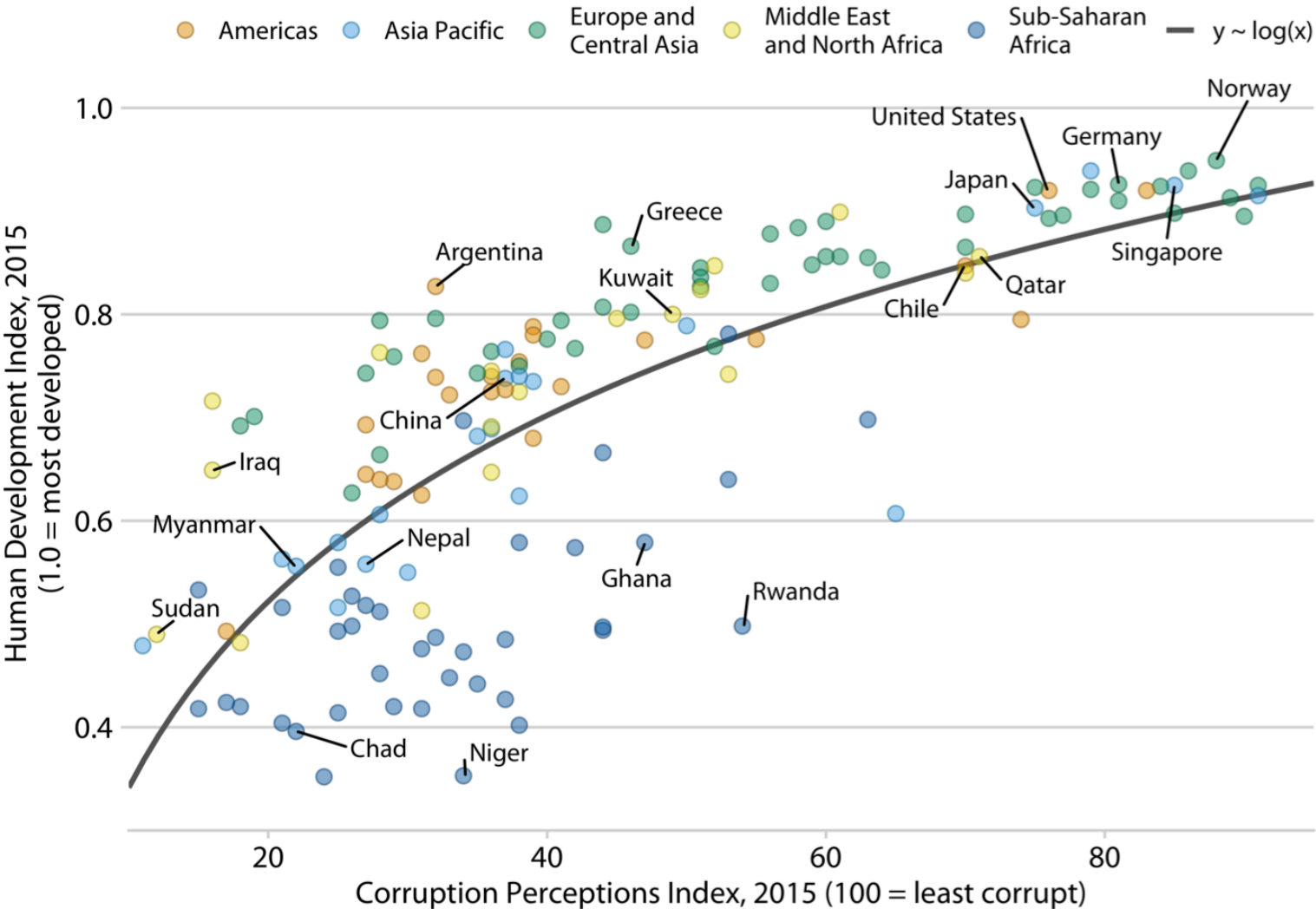
Edzer
Pebesma

What do we need?

- Powerful styling options
- Broad selection of plot types
- **Sophisticated text annotations**
- Plot composition

Corruption and human development

The most developed countries experience the least corruption



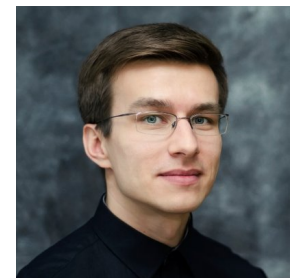
Data sources: Transparency International & UN Human Development Report

ggrepel: Smart annotation of plot elements

```
library(ggrepel)
```

```
mtcars$car <- rownames(mtcars) # add column of car names
```

```
mtcars %>% filter(displ < 200) %>%  
  ggplot(aes(displ, mpg, label = car)) +  
  geom_point() +  
  geom_text_repel()
```



Kamil Slowikowski

ents

ggrep

library(ggplot2)

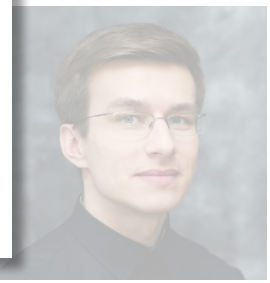
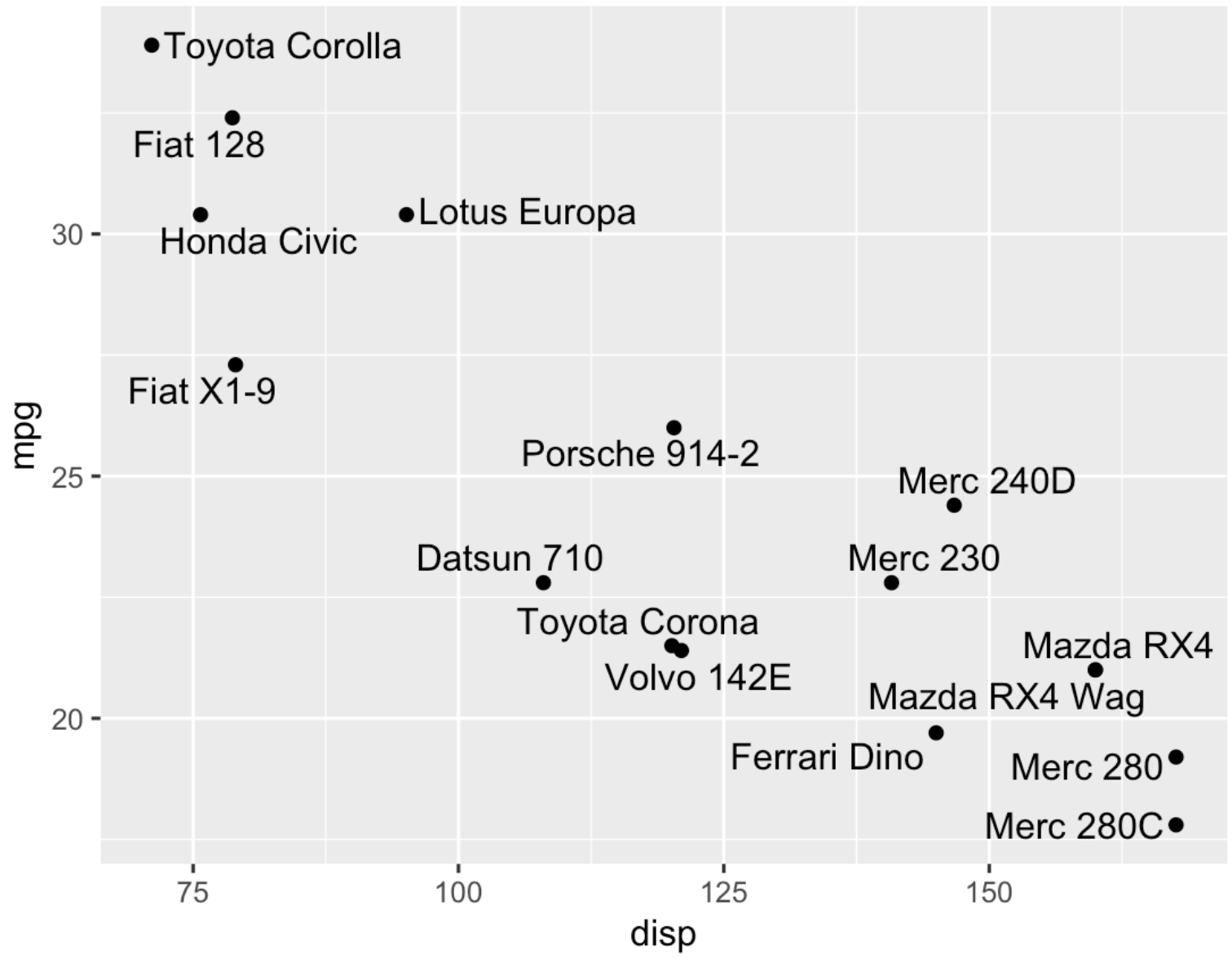
mtcars\$mpg

mtcars %>%

ggplot(aes(x=disp, y=mpg))

geom_point()

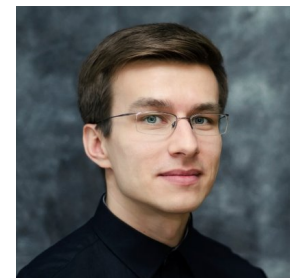
geom_text(aes(x=disp, y=mpg, label=manufacturer))



Kamil Slowikowski

ggrepel: Smart annotation of plot elements

```
mtcars %>% filter(displacement < 200) %>%  
  ggplot(aes(displacement, mpg, label = car)) +  
  geom_point() +  
  ggrepel::geom_text_repel(xlim = c(180, 250), hjust = 0) +  
  xlim(70, 240)
```



Kamil Slowikowski

ggrep

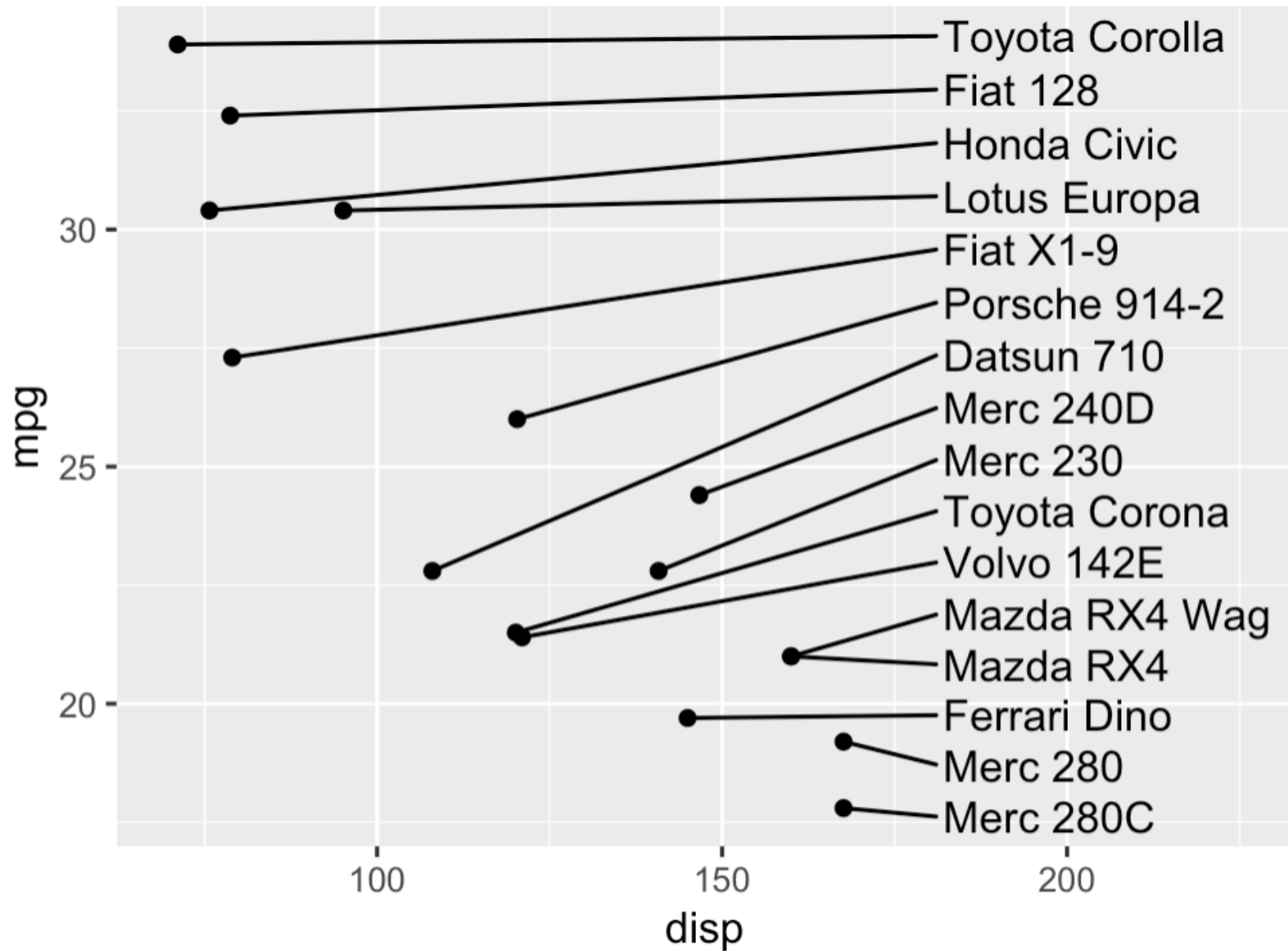
mtcars

ggplot

geom_

geom_

xlim(

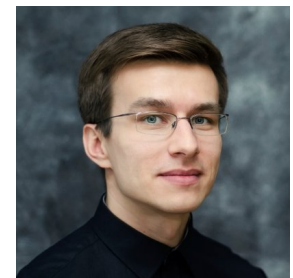


ts



ggrepel: Smart annotation of plot elements

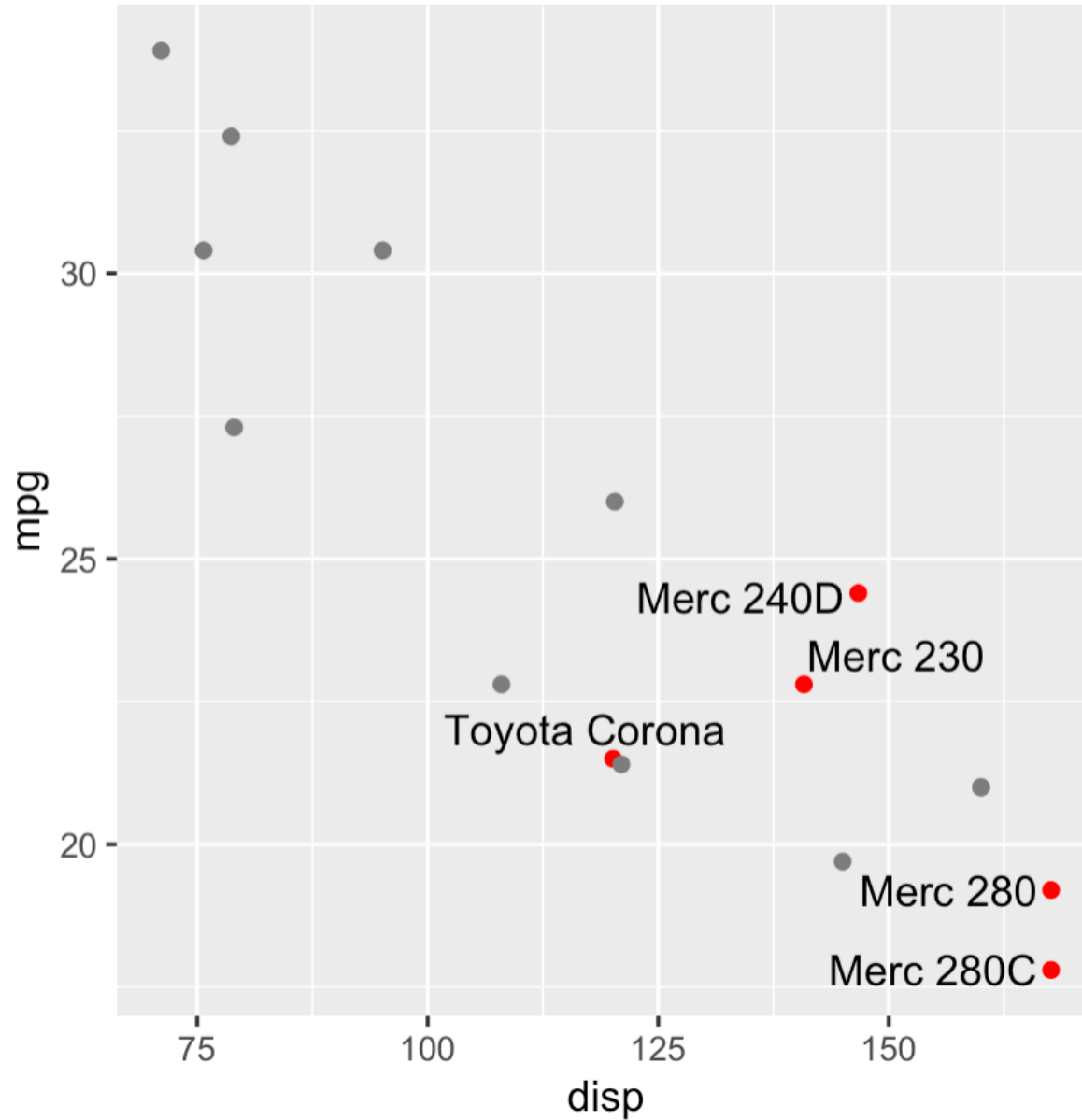
```
mtcars %>% filter(displ < 200) %>%  
  ggplot(aes(displ, mpg)) +  
  geom_point(aes(color = factor(am))) +  
  geom_text_repel(aes(label = ifelse(am == 0, car, ""))) +  
  scale_color_manual(  
    name = "transmission",  
    labels = c("manual", "automatic"),  
    values = c("red", "gray50")  
  )  
)
```



Kamil Slowikowski

ggrep

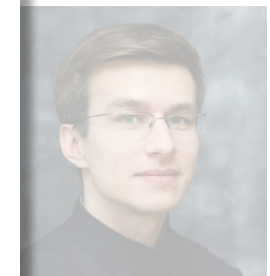
```
mtcars %>%  
  ggplot()  
  geom_point()  
  scale_x_continuous()  
  scale_y_continuous()  
  labs(title = "mtcars",  
        x = "displacement",  
        y = "mpg")  
  theme_minimal()
```



transmission
● manual
● automatic

ents

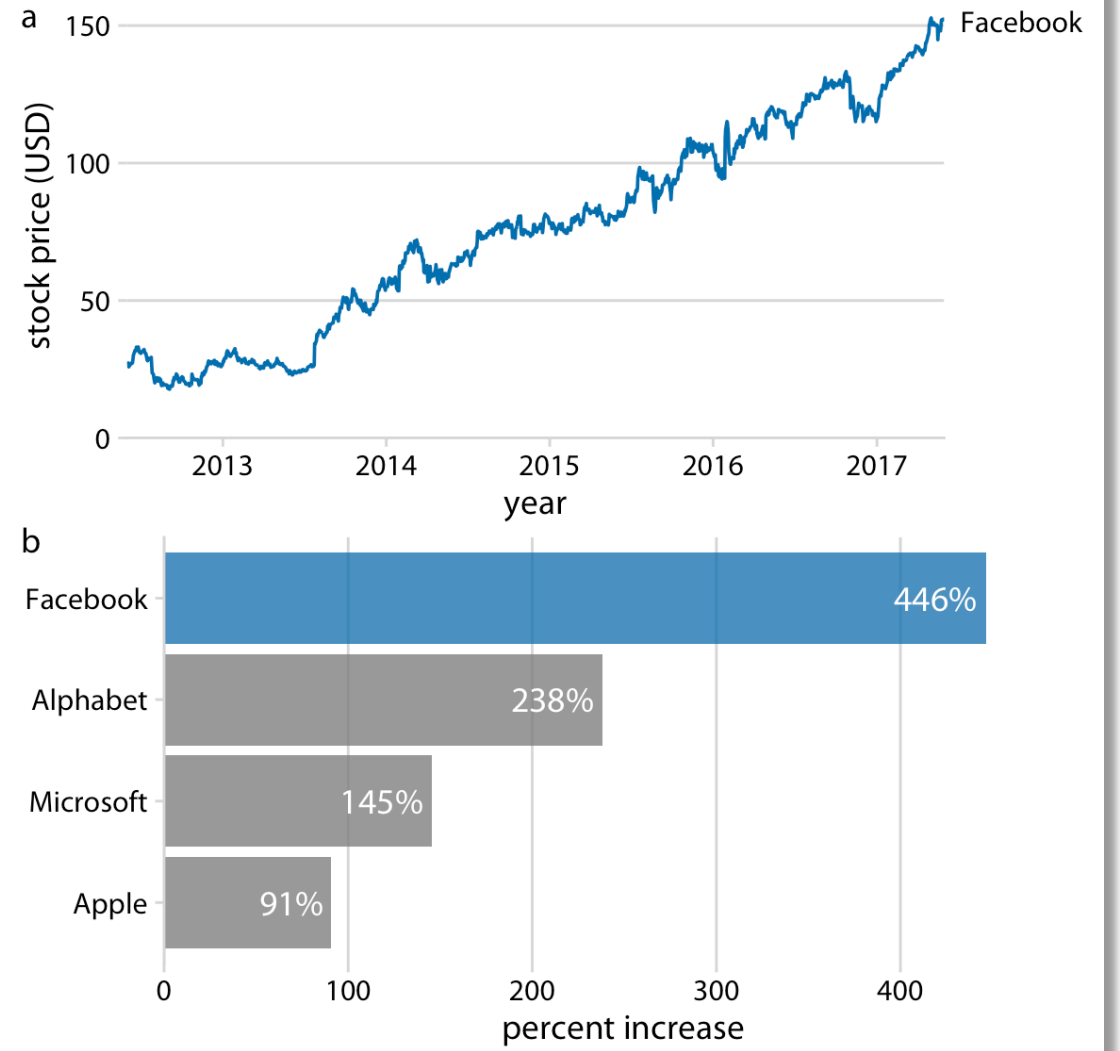
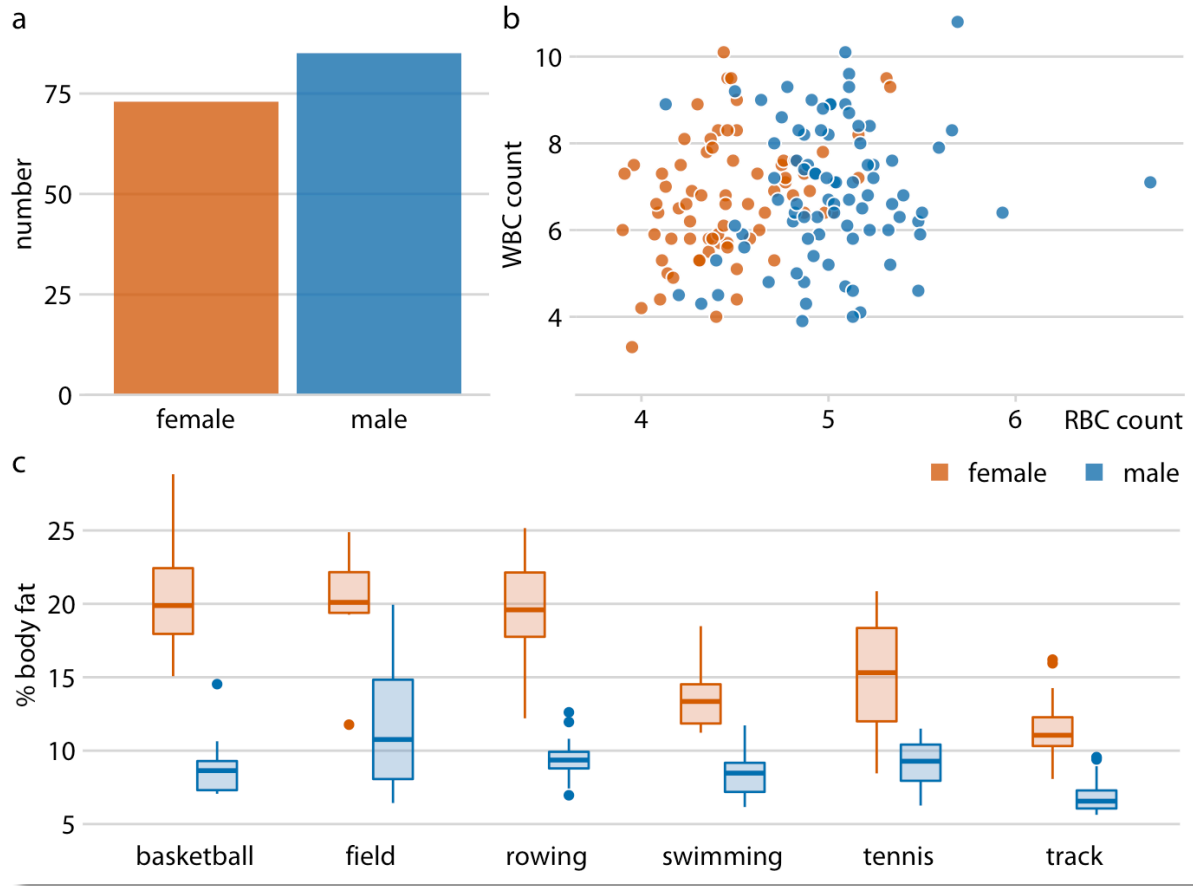
r, "))) +



Kamil Slowikowski

What do we need?

- Powerful styling options
- Broad selection of plot types
- Sophisticated text annotations
- Plot composition



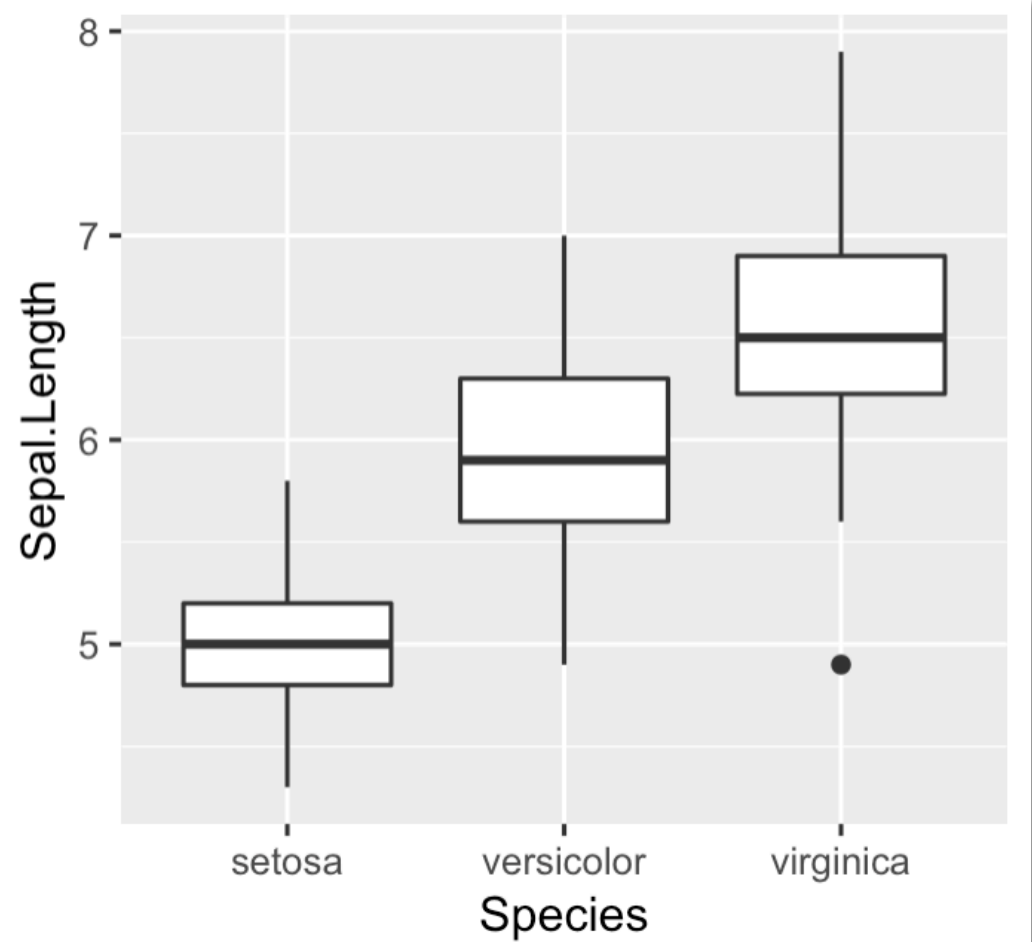
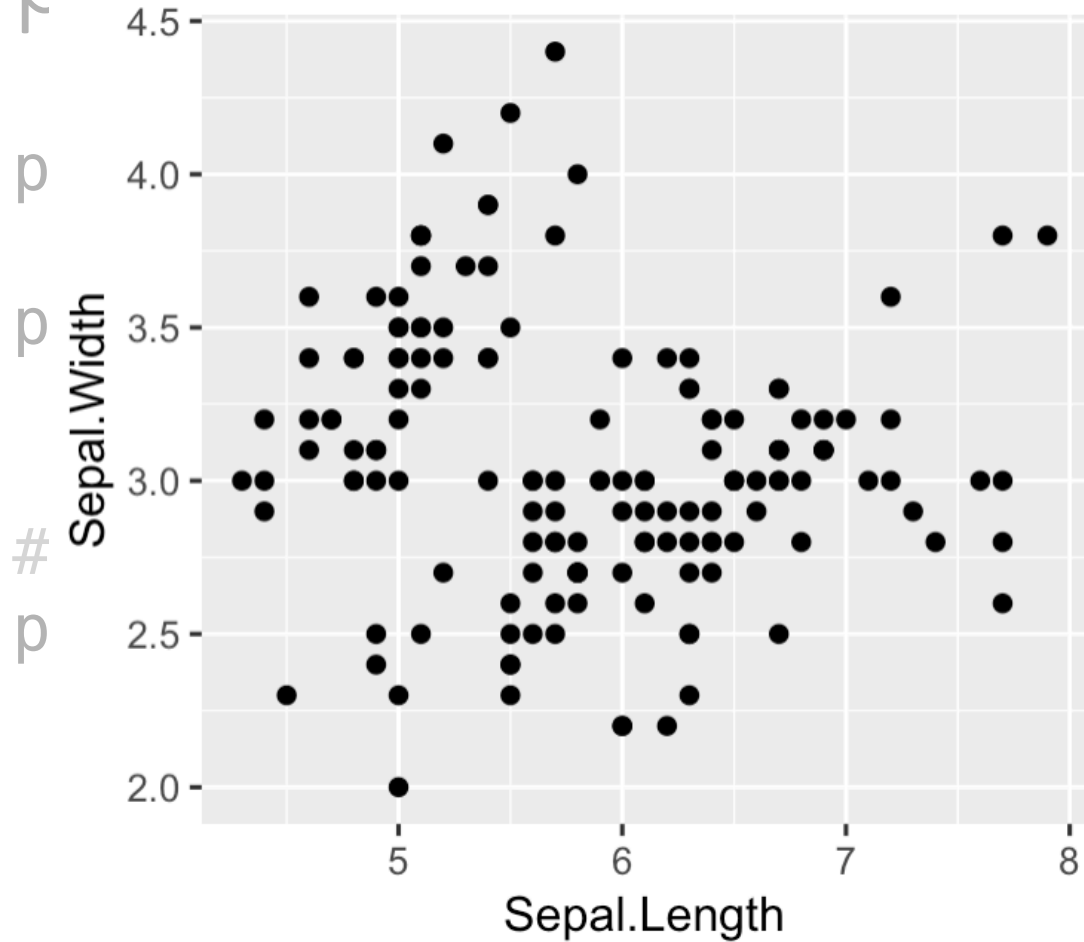
patchwork: A grammar of plot composition

```
p1 <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) +  
      geom_point()  
p2 <- ggplot(iris, aes(Species, Sepal.Length)) +  
      geom_boxplot()  
  
# add plots to place them side-by-side  
p1 + p2
```



Thomas Lin
Pederson

patchwork: Δ grammar of plot composition



Thomas Lin
Pederson

patchwork: A grammar of plot composition

```
p1 <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) +  
      geom_point()  
p2 <- ggplot(iris, aes(Species, Sepal.Length)) +  
      geom_boxplot()  
  
# divide plots to place them on top of each other  
p1 / p2
```

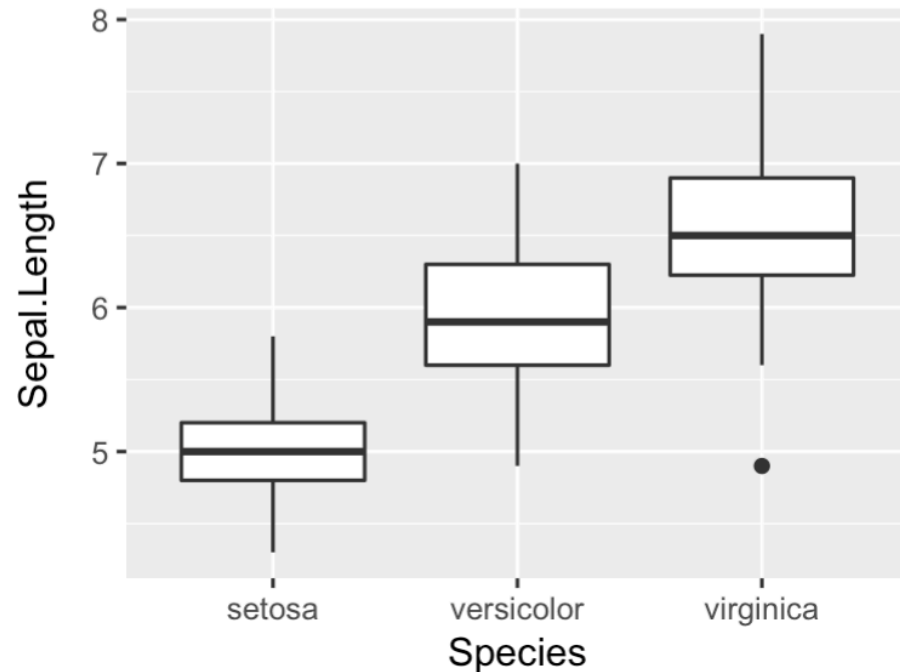
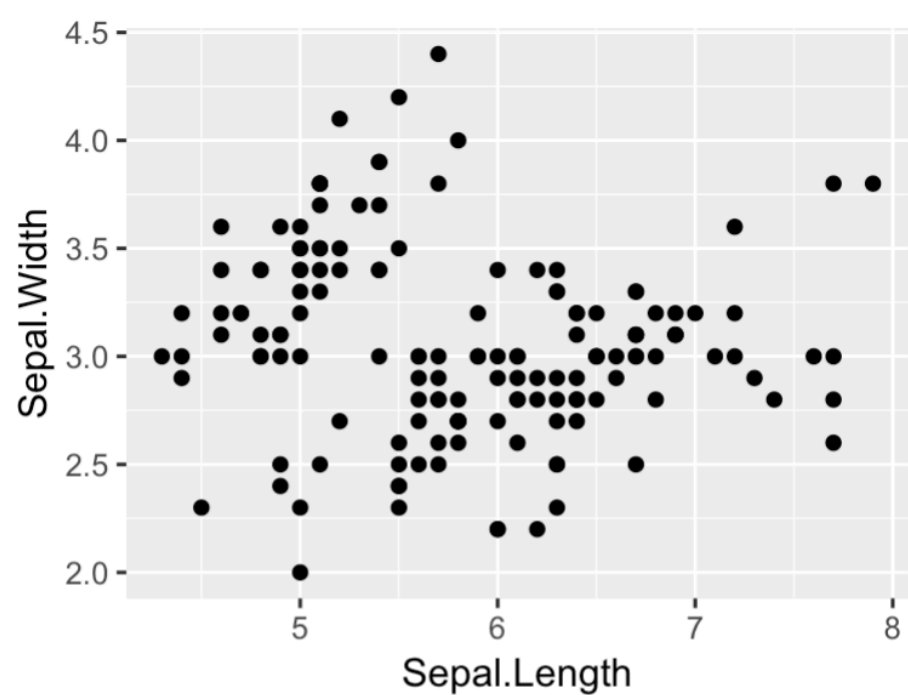


Thomas Lin
Pederson

patchwork: A

```
p1 <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) +  
  geom_point()  
p2 <- ggplot(iris, aes(Sepal.Length)) +  
  geom_boxplot()
```

```
# divide plots to be viewed side-by-side  
p1 / p2
```



composition

```
aes(Sepal.Length, Sepal.Width)) +  
  geom_point()  
  aes(Sepal.Length)) +  
  geom_boxplot()
```

each other



Thomas Lin
Pederson

patchwork: A grammar of plot composition

```
p3 <- ggplot(iris, aes(Sepal.Length)) +  
      geom_density(fill = "gray60") +  
      facet_wrap(~Species)
```

```
# nested arrangements are possible  
p3 / (p1 + p2)
```

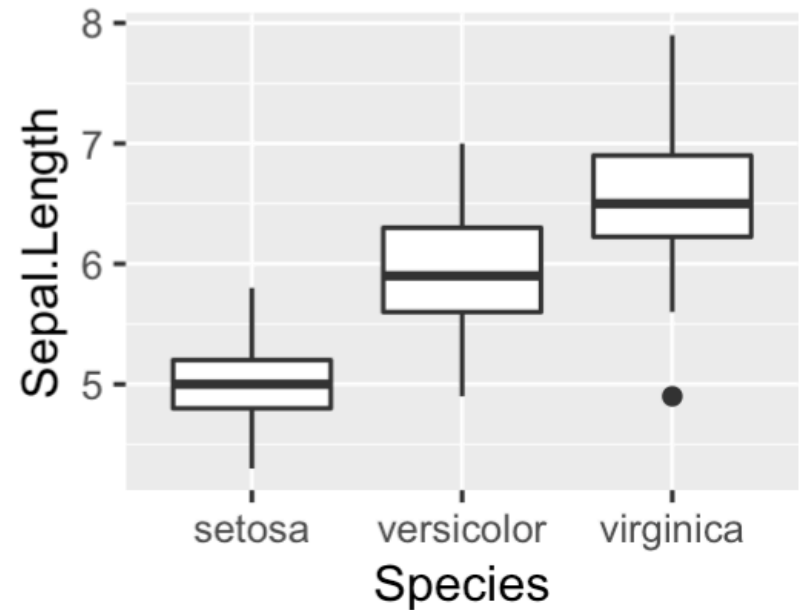
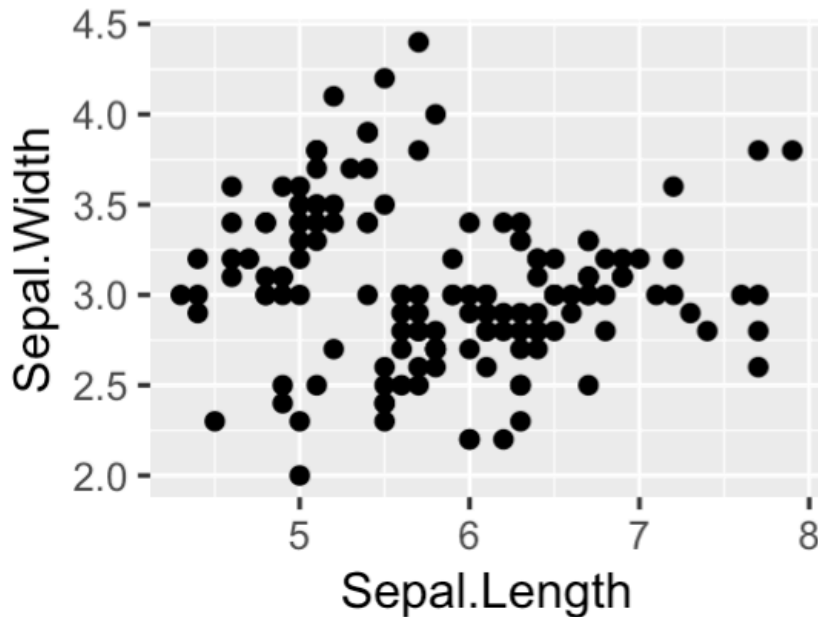
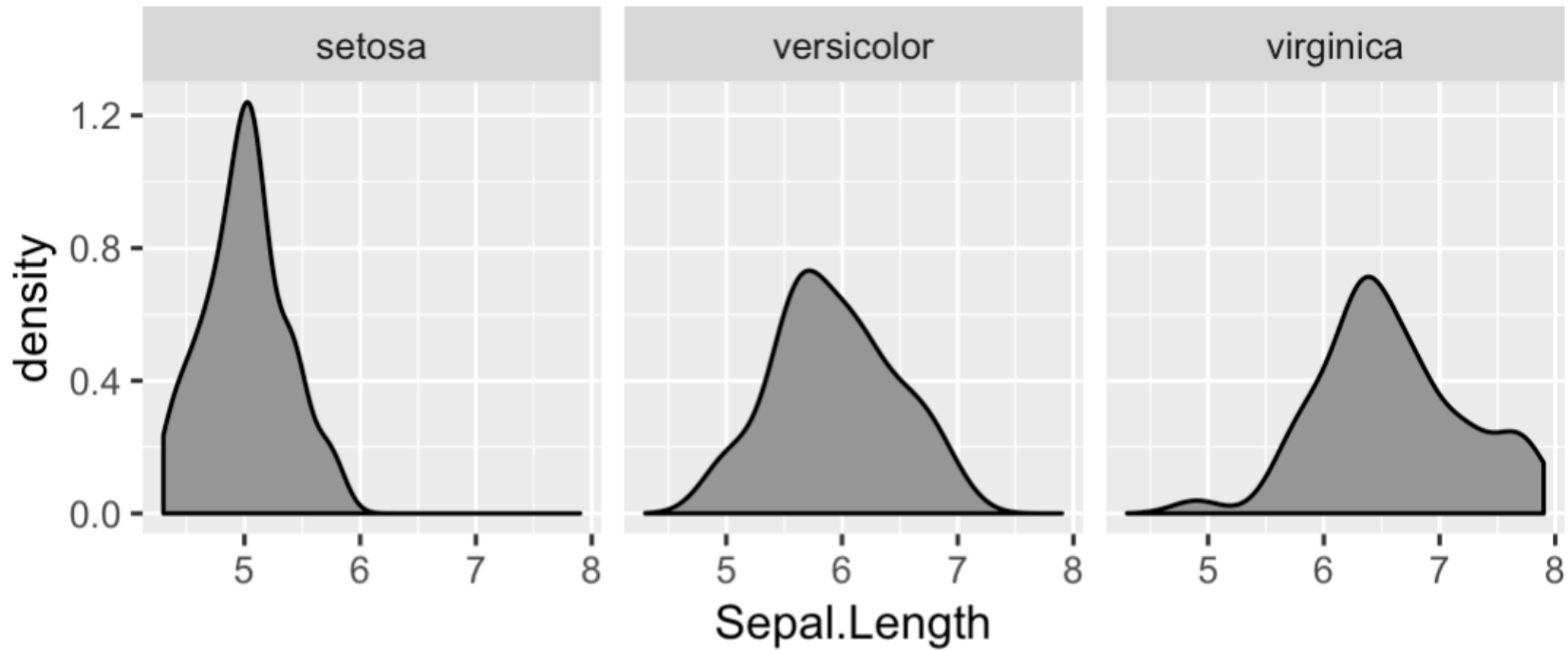


Thomas Lin
Pederson

patch

```
p3 <- g
```

```
# neste  
p3 / (p
```



patchwork: A grammar of plot composition

```
p3 <- ggplot(iris, aes(Sepal.Length)) +  
      geom_density(fill = "gray60") +  
      facet_wrap(~Species)
```

```
# apply theme recursively to all plots  
p3 / (p1 + p2) & theme_minimal()
```

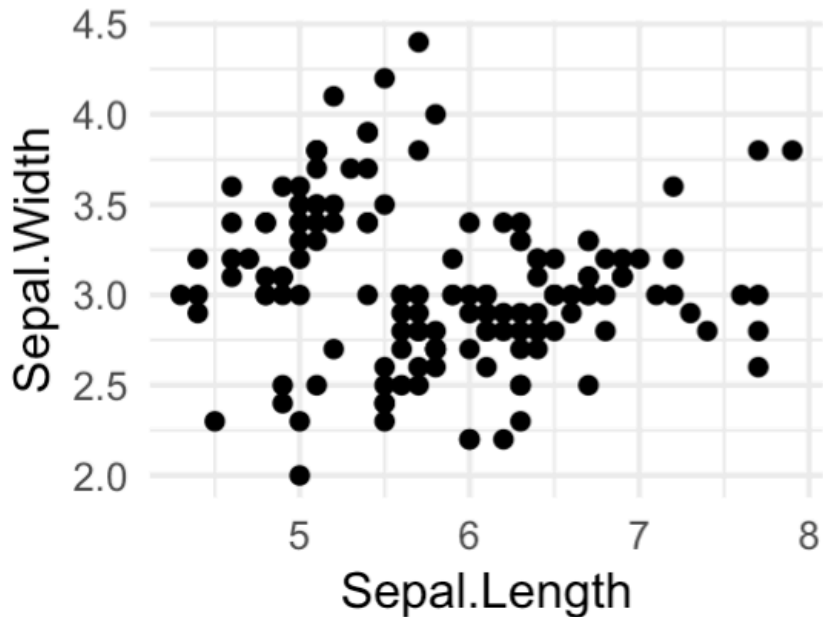
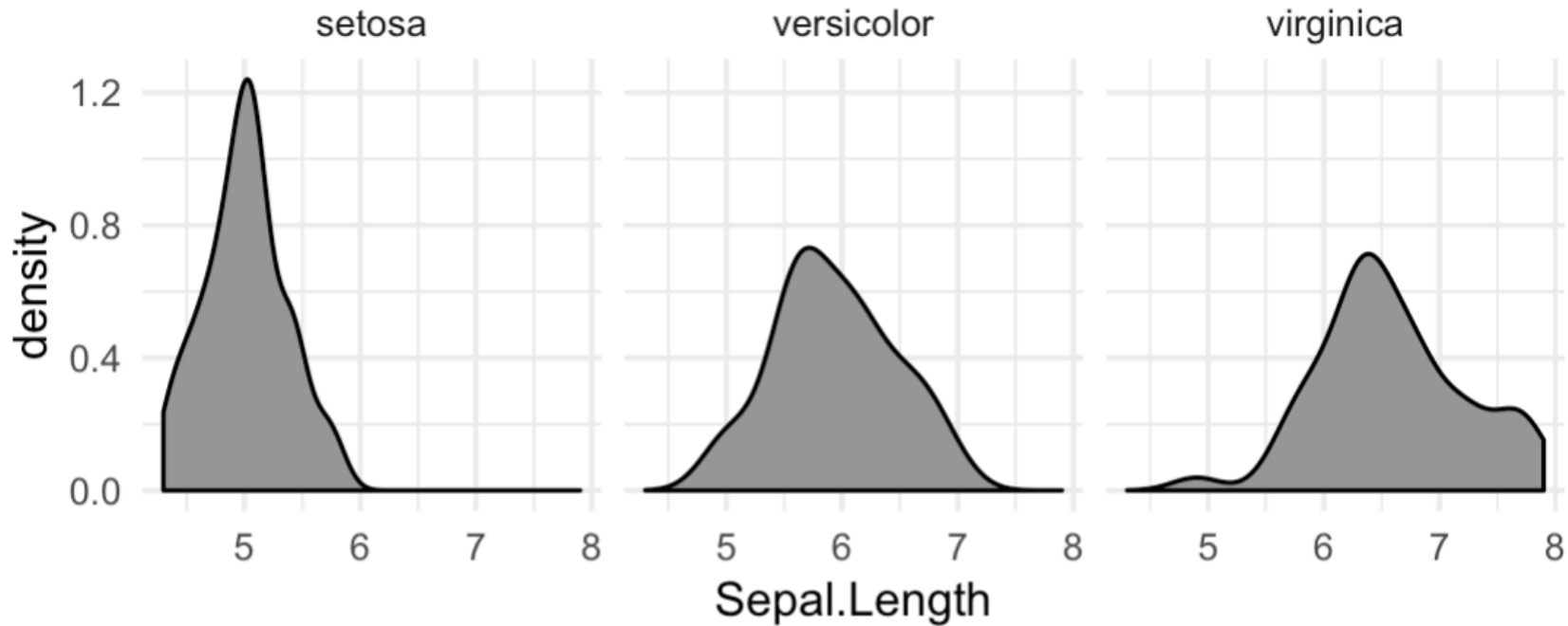


Thomas Lin
Pederson

patch

```
p3 <- g
```

```
# apply  
p3 / (p
```



Read the source, Luke

tidyverse / ggplot2

Unwatch 330 Star 3,821 Fork 1,426

Code Issues 152 Pull requests 26 Actions Wiki Security Insights

An implementation of the Grammar of Graphics in R <https://ggplot2.tidyverse.org>

r visualisation data-visualisation

4,291 commits 15 branches 26 releases 1 environment 184 contributors GPL-2.0

Branch: master New pull request Create new file Upload files Find File Clone or download

clauswilke Allow empty annotations. (#3320) Latest commit 1f6f0cb 15 days ago

.github	Move CODE_OF_CONDUCT.md to the main directory (#2973)	7 months ago
R	Allow empty annotations. (#3320)	15 days ago
data-raw	Update economics data (#2962)	3 months ago
data	Update economics data (#2962)	3 months ago
icons	Tweak icons	a year ago

Acknowledgments

ggplot2 team

Hadley Wickham

Winston Chang

Lionel Henry

Thomas Lin Pedersen

Kohske Takahashi

Hiroaki Yutani

Kara Woo

Specific package authors

Edzer Pebesma (sf)

Kamil Slowikowski (ggrepel)

Thomas Lin Pedersen (patchwork)

The broader R community

Mara Averick

Yihui Xie

RStudio

many others

Claus O. Wilke receives funding from: UT Austin, R Consortium, NIH, NSF