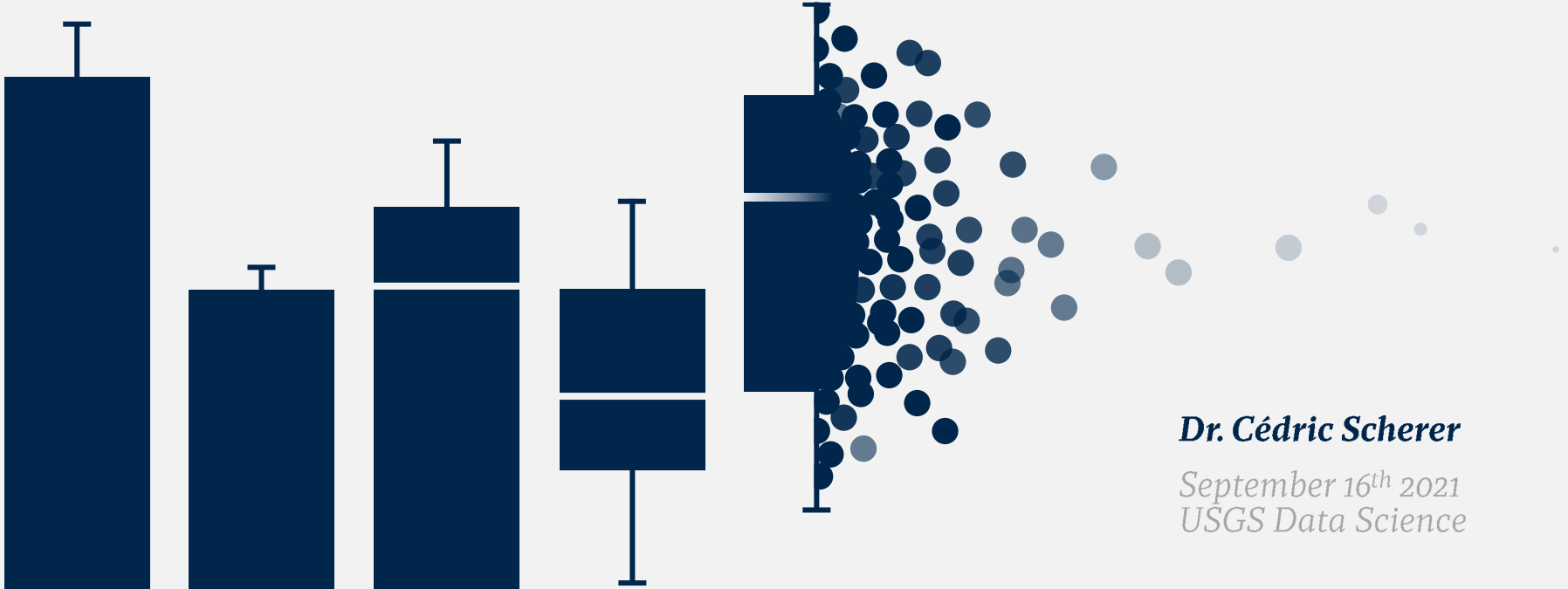
BEYOND BAR AND BOX PLOTS

Chart alternatives and how to design them with ggplot2



Dr. Cédric Scherer

September 16th 2021
USGS Data Science

Cédric Scherer

Freelance Data Visualization Specialist
Computational Ecologist at IZW Berlin

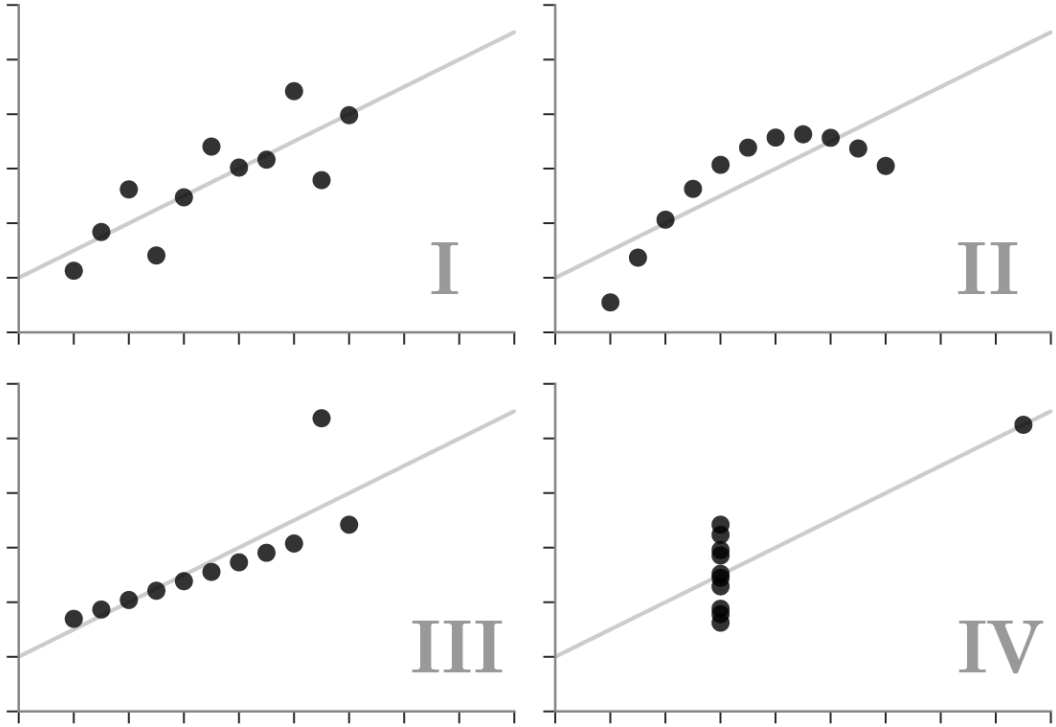
cedricscherer.com @CedScherer @z3tt @cedscherer

...make ***both calculations and graphs****.*
Both sorts of output should be studied;
each will ***contribute to understanding****.*

F. J. Anscombe (1973)

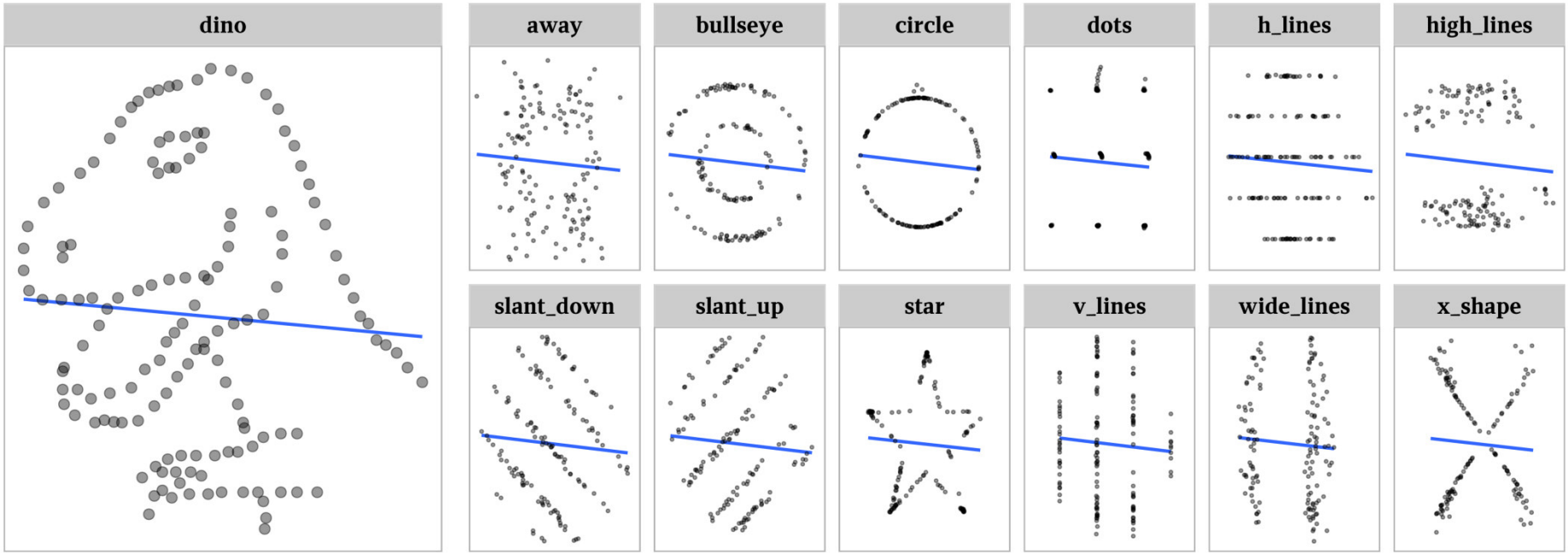
Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



Matejka & Fitzmaurice (2017)

The Datasaurus by Alberto Cairo shows us why visualisation is important, not just summary statistics.



PLOS | BIOLOGY

PERSPECTIVE

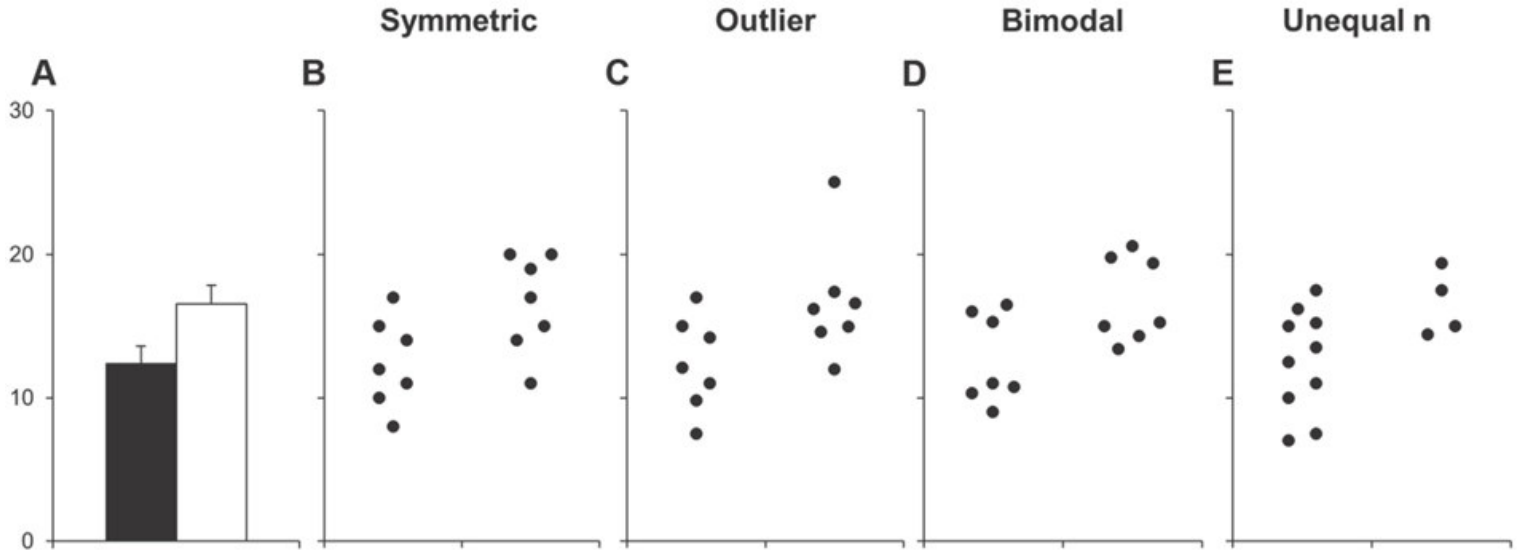
Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

Tracey L. Weissgerber[1]*, Natasa M. Milic[1,2], Stacey J. Winham[3], Vesna D. Garovic[1]

1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America, **2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States of America

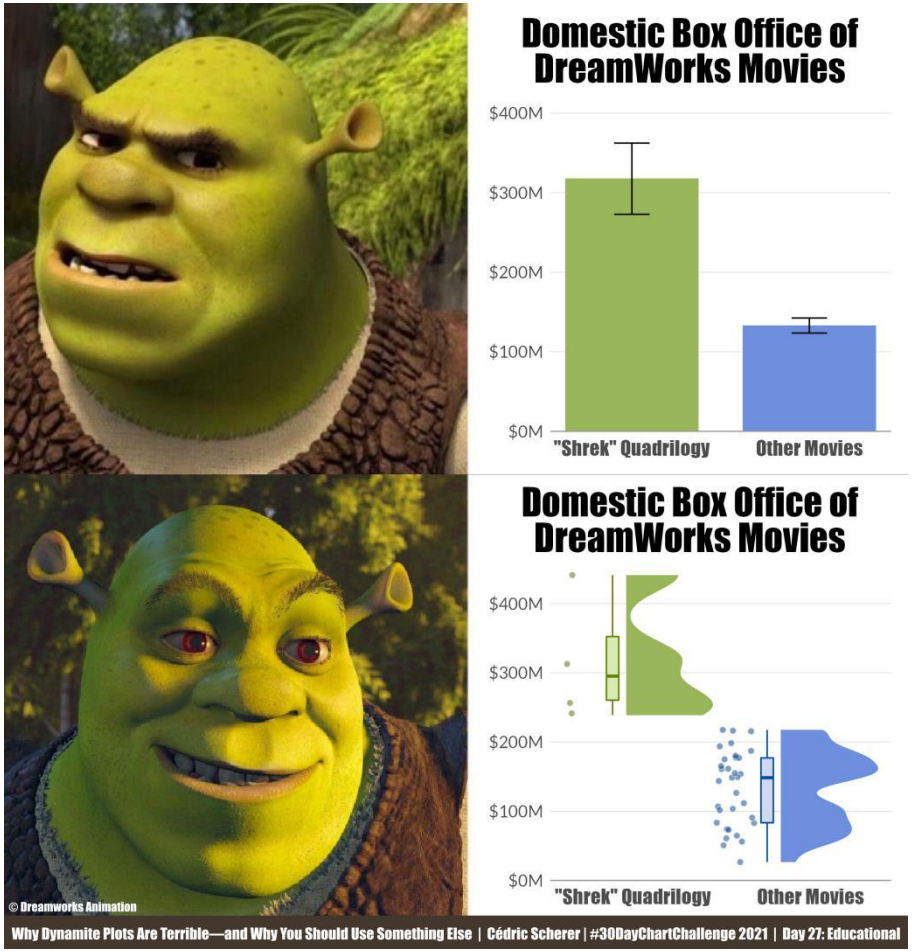
* weissgerber.tracey@mayo.edu

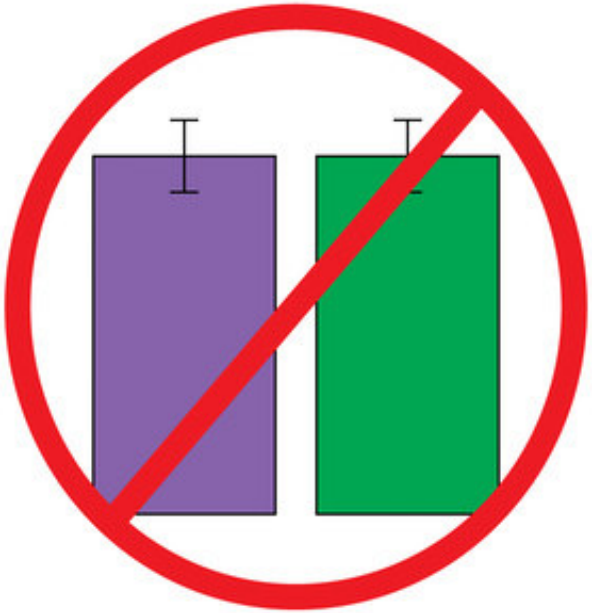
Weissgerber et al. (2015) PLoS Biol 13:e1002128 (10.1371/journal.pbio.1002128)



Test				
	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

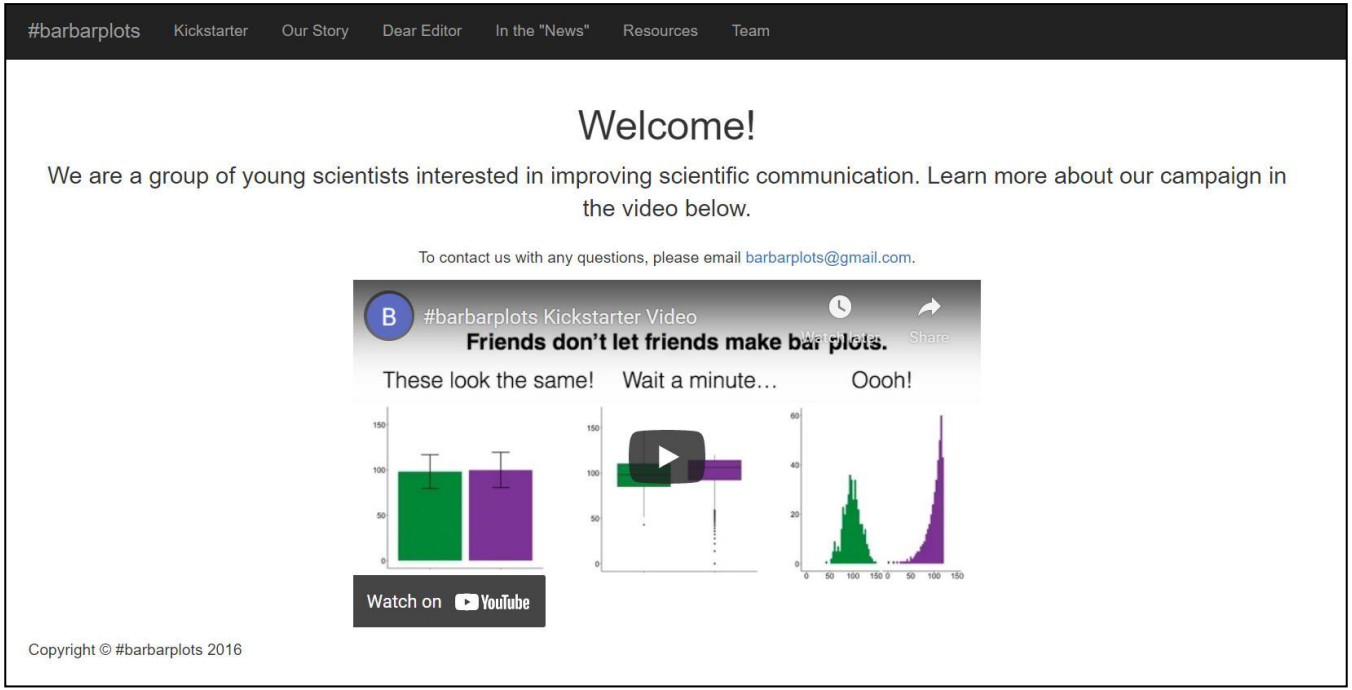
Weissgerber et al. (2015) PLoS Biol 13:e1002128 (10.1371/journal.pbio.1002128)


Domestic Box Office of DreamWorks Movies
$400M
$300M
$200M
$100M
$0M
"Shrek" Quadrilogy
Other Movies
Domestic Box Office of DreamWorks Movies
$400M
$300M
$200M
$100M
$0M
"Shrek" Quadrilogy
Other Movies
© Dreamworks Animation
Why Dynamite Plots Are Terrible—and Why You Should Use Something Else | Cédric Scherer | #30DayChartChallenge 2021 | Day 27: Educational



#barbarplots

barbarplots.github.io



barbarplots.github.io
youtu.be/LFDbqw2xPbQ

Dear Editor

Data visualization is a complex topic in the experimental sciences. While there are many ways to display data, many researchers choose to use bar plots. Generally, these plots only depict a group mean and standard error (or deviation). Unfortunately, most data are not as clean as bar plots make them seem, and since bar plots reveal very little about the distribution of the data, this kind of visualization can be misleading [1,2,3]. A further issue is that of the bar itself, which implies that the base of the y-axis is meaningful, which is not necessarily the case. The bar can then mislead readers [3].

We are a group of young scientists who recently started an initiative called “#barbarplots” aimed at raising awareness and improving scientific communication. As part of this campaign, 169 individuals pledged over 3,400€ in a Kickstarter project to send this package to editors of top journals. Journal editors like you have the power to greatly influence trends in the field. We would like to suggest that you begin a conversation with your editorial board about your journal’s stance on data visualization, and whether you wish to encourage authors to use more informative techniques to plot distributions of data. Many high impact journals, such as Nature Neuroscience, the Journal of Neuroscience, and PLoS Biology, have already taken the step to #barbarplots in an effort to make articles more transparent.

To better spread the word, we wish to encourage your participation in our social media campaign. A t-shirt is enclosed in this package: put it on, take a selfie, and share it on Twitter or Facebook with the hashtag #barbarplots.

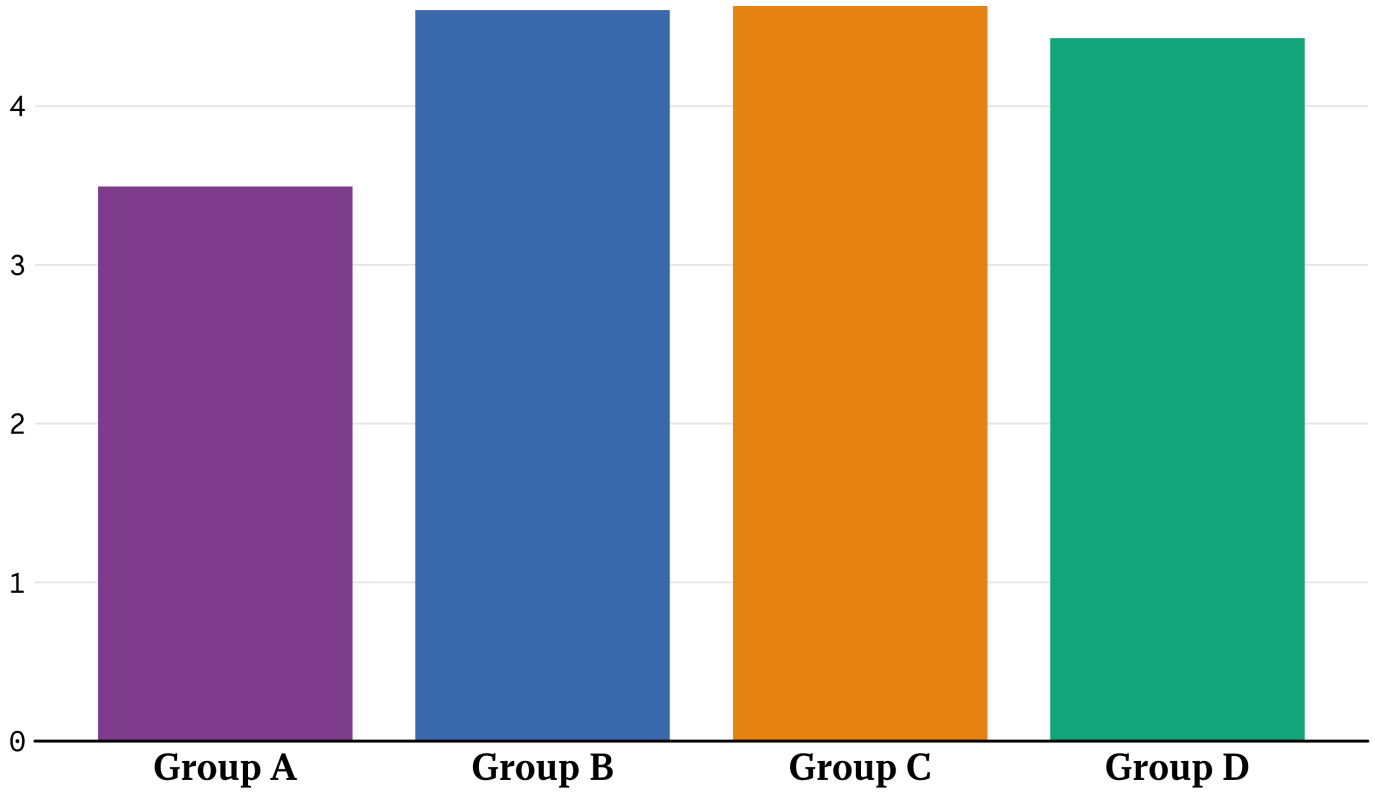
We hope that this initiative will encourage discussions between colleagues on the merits of various data visualization techniques and on transparency in science in general.

References

1. Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biology, 13*(4), 1–10. doi:10.1371/journal.pbio.1002128
2. Weissgerber, T. L., Garovic, V. D., Savic, M., Winham, S. J., & Milic, N. M. (2016). From Static to Interactive: Transforming Data Visualization to Improve Transparency. *PLoS Biology, 14*(6), e1002484. doi:10.1371/journal.pbio.1002484
3. Saxon, E. (2015). Beyond bar charts. *BMC Biology, 13*(1), 1–2. doi:10.1186/s12915-015-0169-6

BAR PLOT

geom_bar(fun = "summary")



CODES

z3tt.github.io/beyond-bar-and-box-plots

GGPLOT2 SETUP

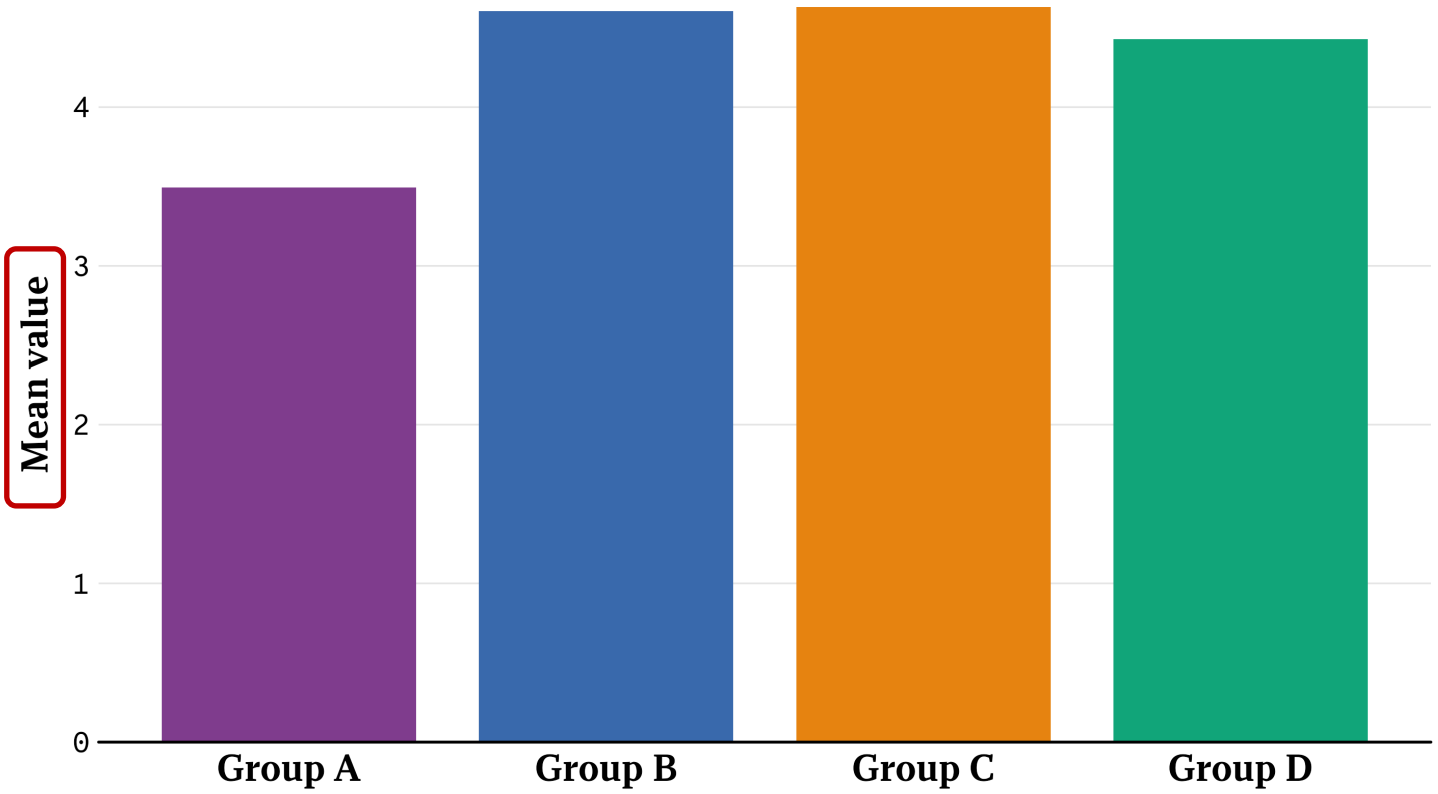
```
## general theme
theme_set(theme_void(base_family = "Roboto"))

theme_update(
  axis.text.x = element_text(color = "black", face = "bold", size = 26,
                             margin = margin(t = 6)),
  axis.text.y = element_text(color = "black", size = 22, hjust = 1,
                             margin = margin(r = 6), family = "Roboto Mono"),
  panel.grid.major.y = element_line(color = "grey90", size = .6),
  axis.line.x = element_line(color = "black", size = 1),
  plot.margin = margin(rep(20, 4))
)

## custom colors
my_pal <- rcartocolor::carto_pal(n = 8, name = "Bold")[c(1, 3, 7, 2)]
```

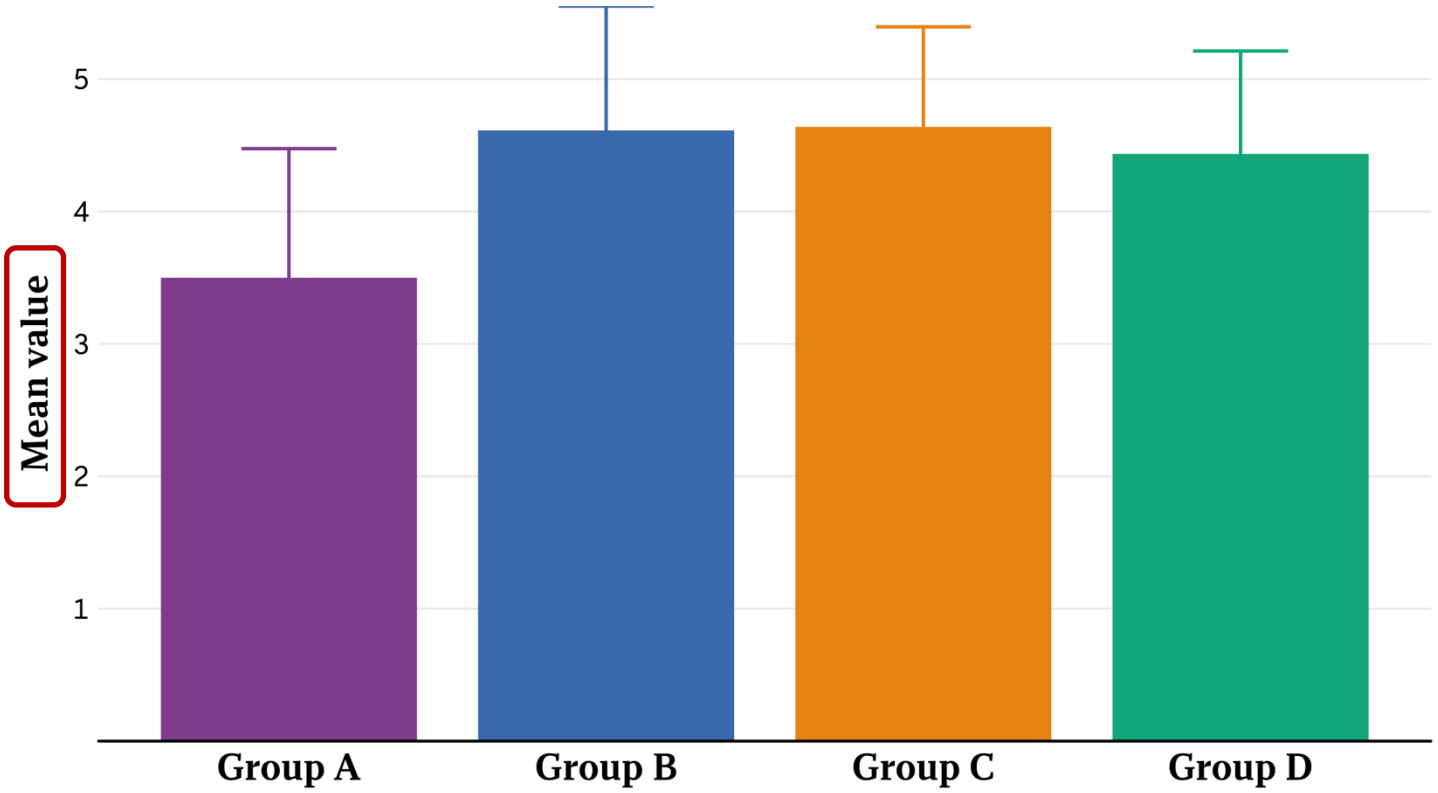
~~BAR PLOT~~

geom_bar(stat = "summary")

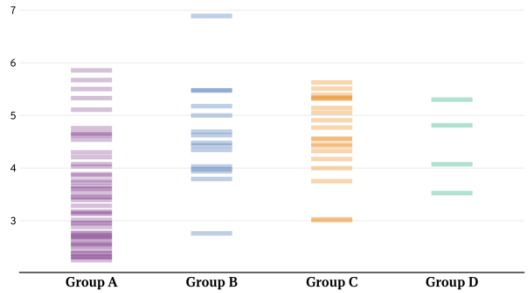


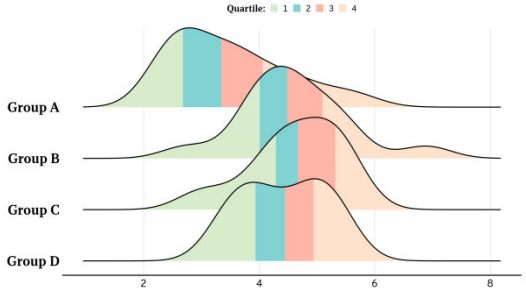
~~DYNAMITE PLOT~~

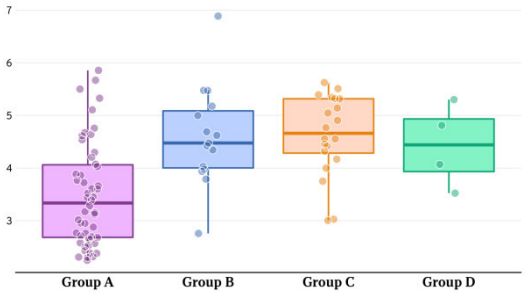
```
geom_bar(stat = "summary") + stat_summary(geom = "errorbar")
```

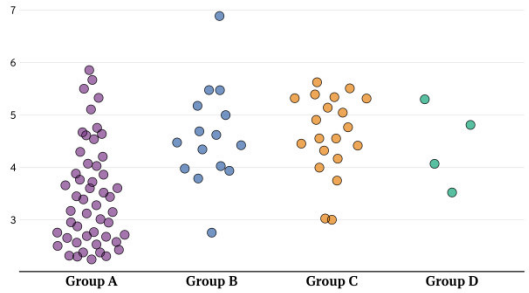


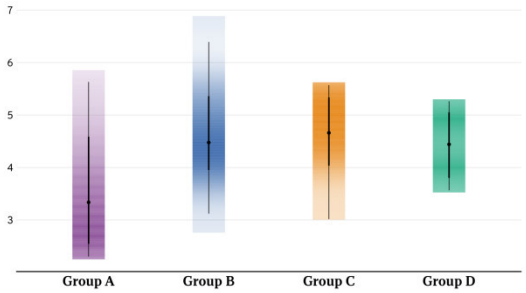

5
4
3
2
1
Group A
Group D

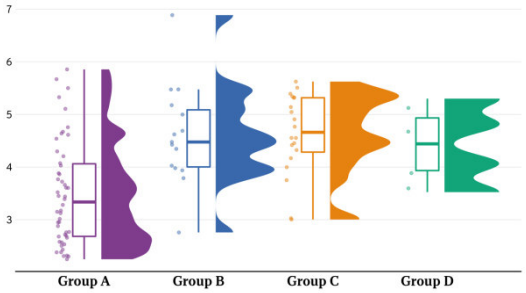

7
6
5
4
3
Group A
Group B
Group C
Group D


Quartile: 1 2 3 4
Group A
Group B
Group C
Group D
2
4
6
8


7
6
5
4
3
Group A
Group B
Group C
Group D

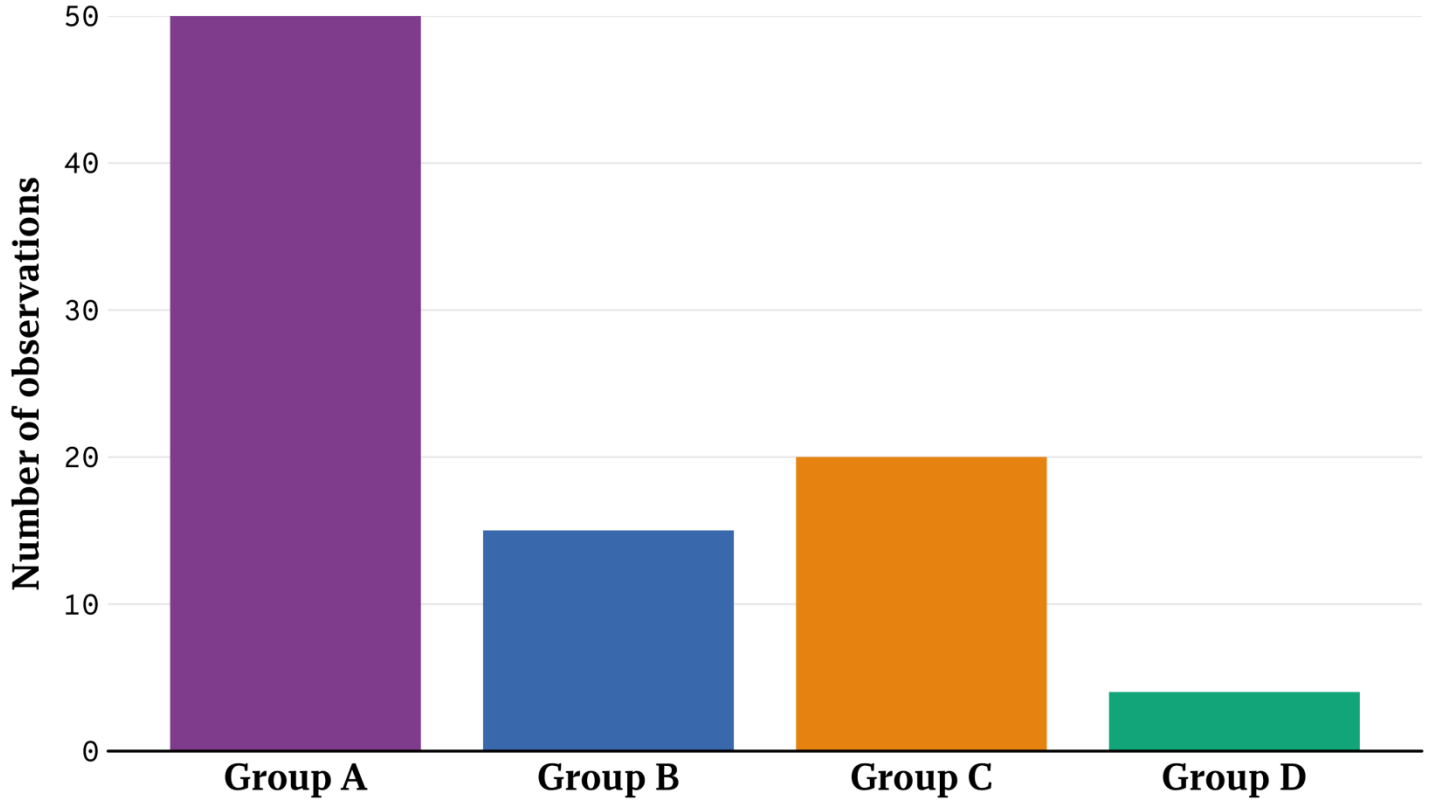

7
6
5
4
3
Group A
Group B
Group C
Group D


7
6
5
4
3
Group A
Group B
Group C
Group D

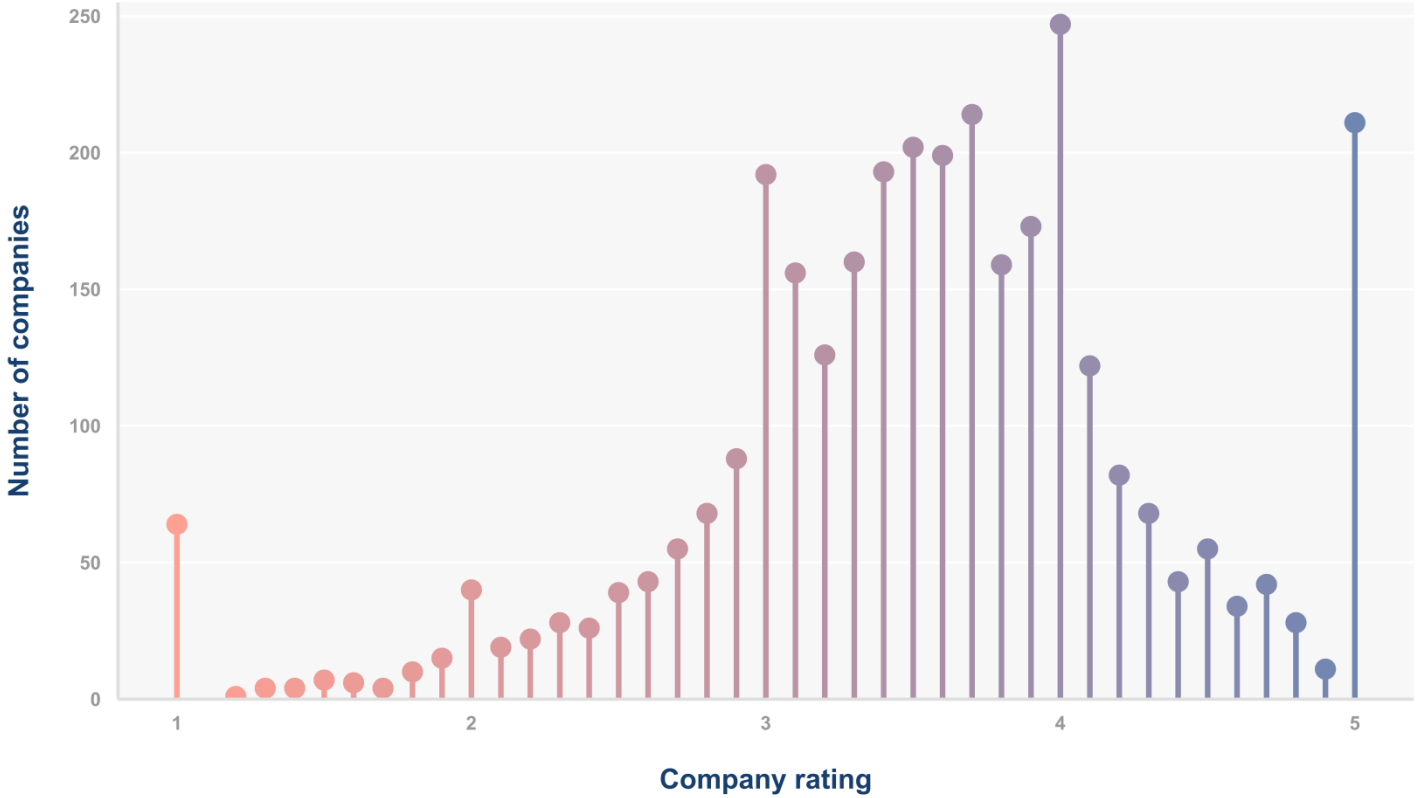

7
6
5
4
3
Group A
Group B
Group C
Group D

BAR PLOT ✓

geom_bar(stat = "count")

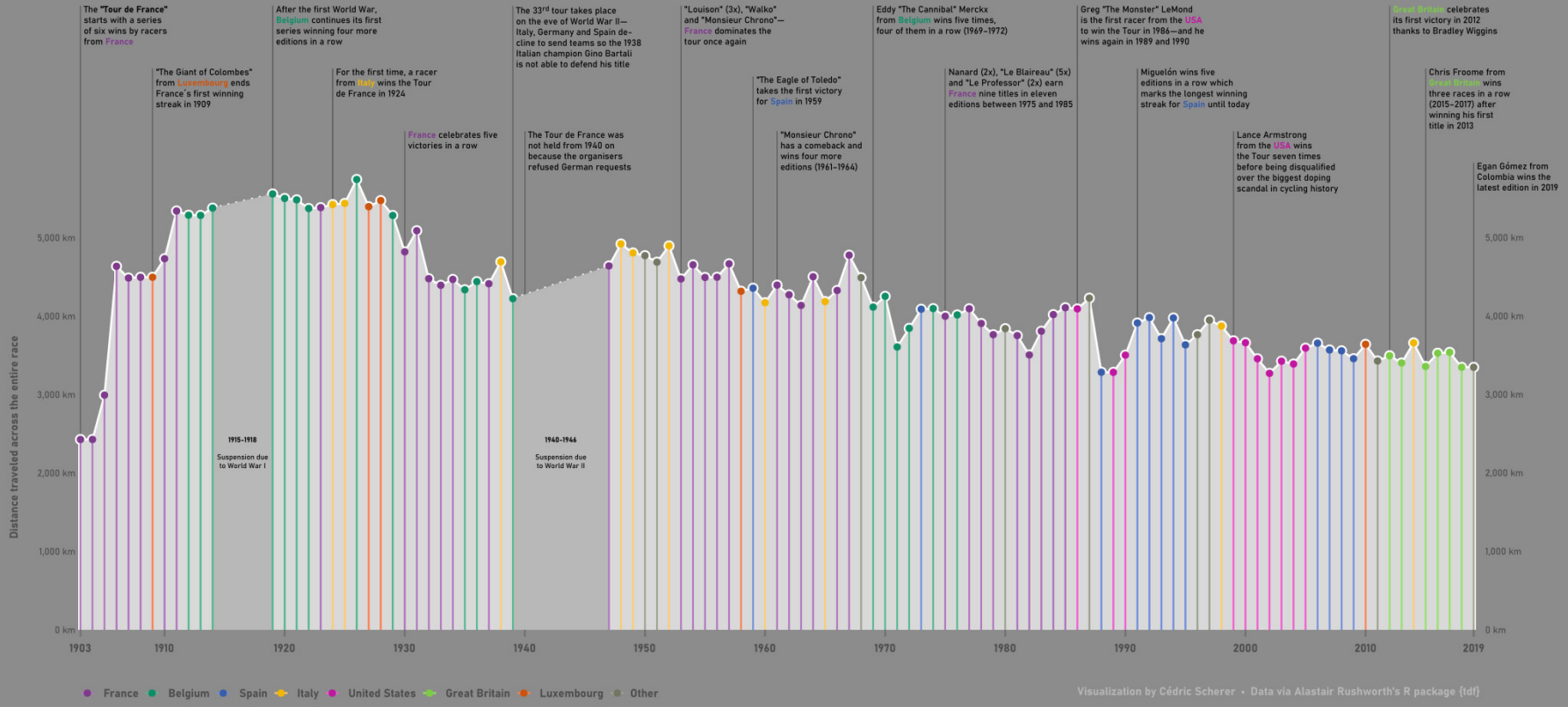


Do more low- or high-rated companies post e-commerce jobs?



Source: Clerk.io

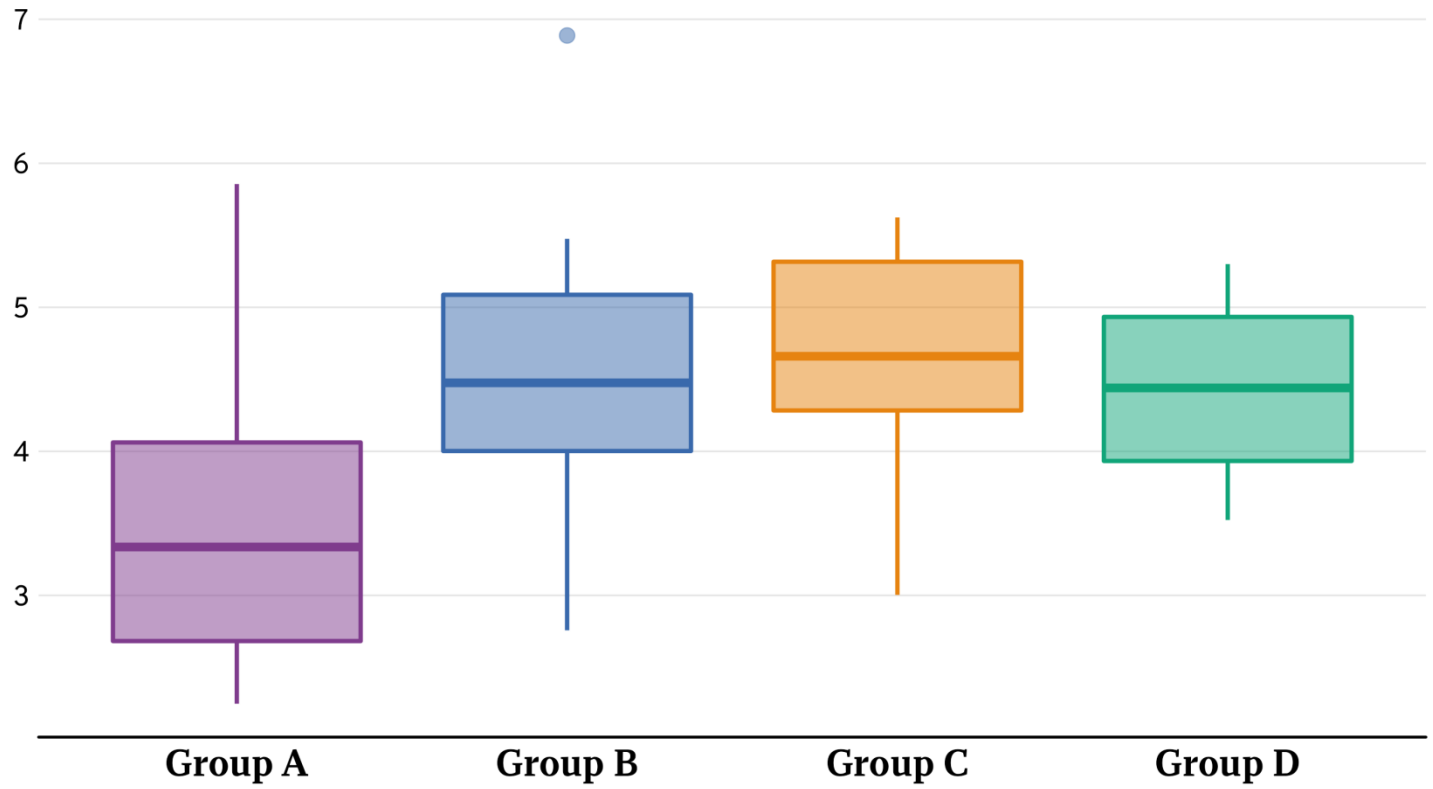
L'histoire du Tour de France de 1903 à 2019



Visualization by Cédric Scherer • Data via Alastair Rushworth's R package {tdf}

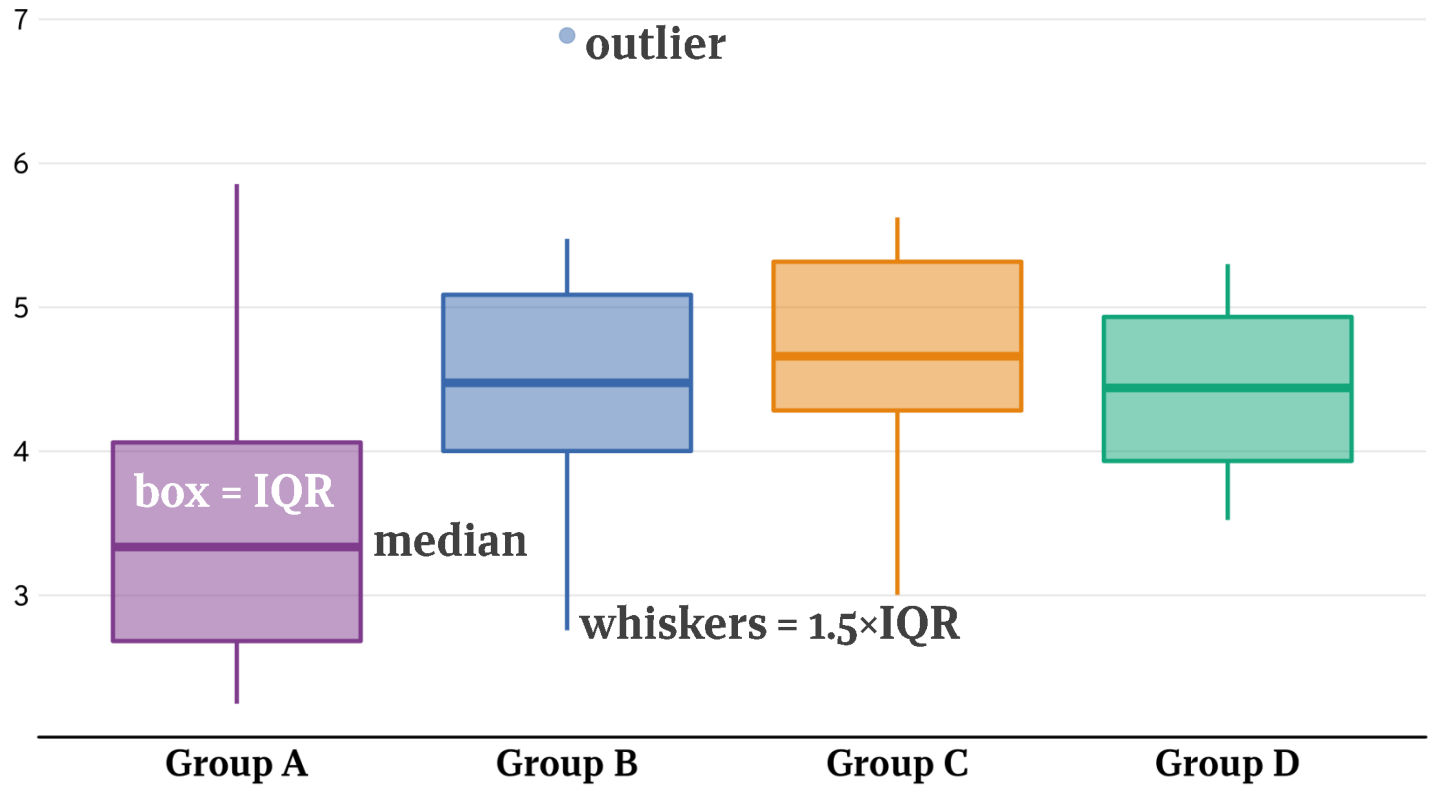
BOX-AND-WHISKERS PLOT

geom_boxplot()



BOX-AND-WHISKERS PLOT

geom_boxplot()



BOX-AND-WHISKERS PLOT

geom_boxplot()

```
g <- ggplot(data, aes(x = group, y = value, color = group, fill = group)) +
  scale_y_continuous(breaks = 1:9) +
  scale_color_manual(values = my_pal, guide = "none") +
  scale_fill_manual(values = my_pal, guide = "none")
```

BOX-AND-WHISKERS PLOT

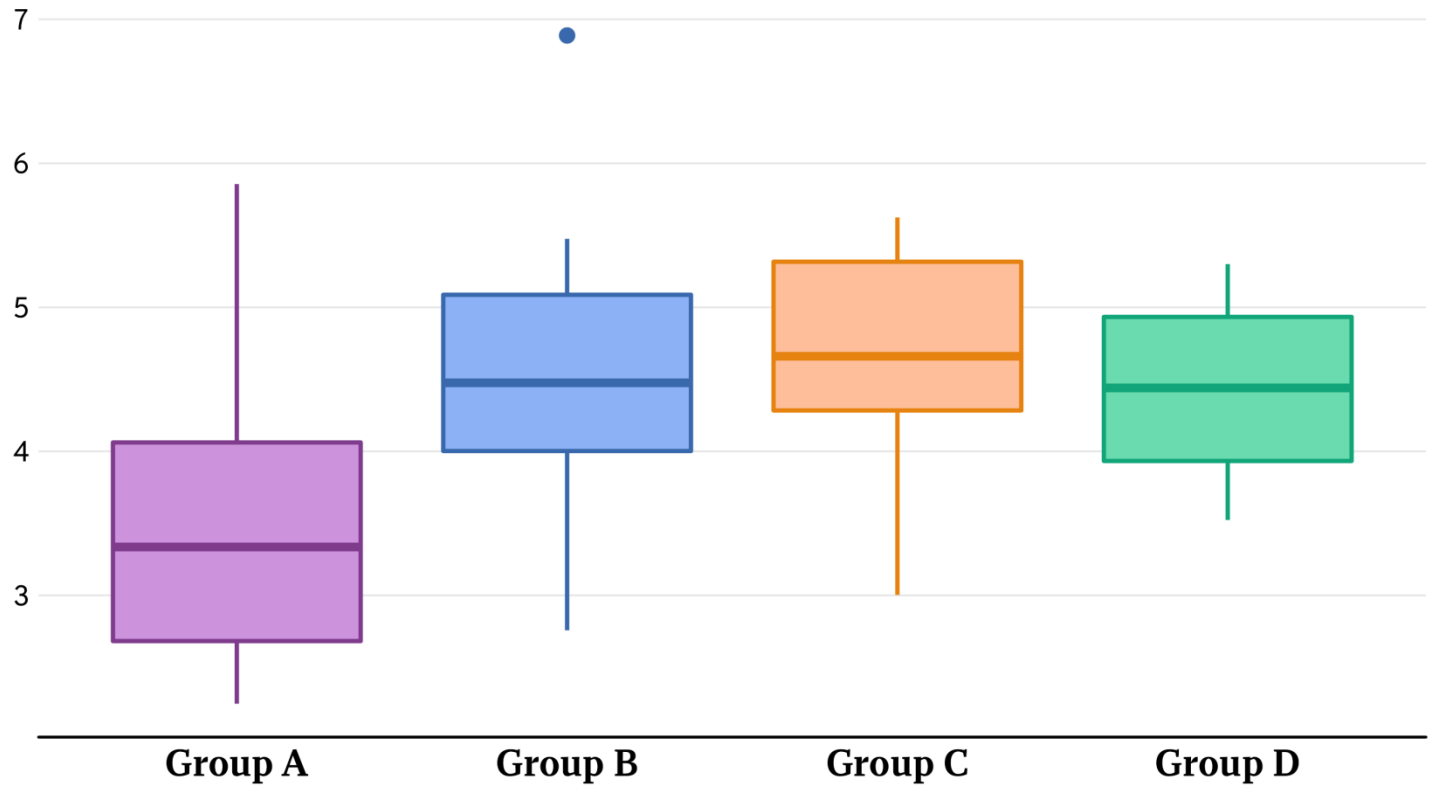
geom_boxplot()

```
g <- ggplot(data, aes(x = group, y = value, color = group, fill = group)) +
  scale_y_continuous(breaks = 1:9) +
  scale_color_manual(values = my_pal, guide = "none") +
  scale_fill_manual(values = my_pal, guide = "none")
```

```
g +
  geom_boxplot(
    aes(fill = group), alpha = .5,
    size = 1.5, outlier.size = 5
  )
```

BOX-AND-WHISKERS PLOT

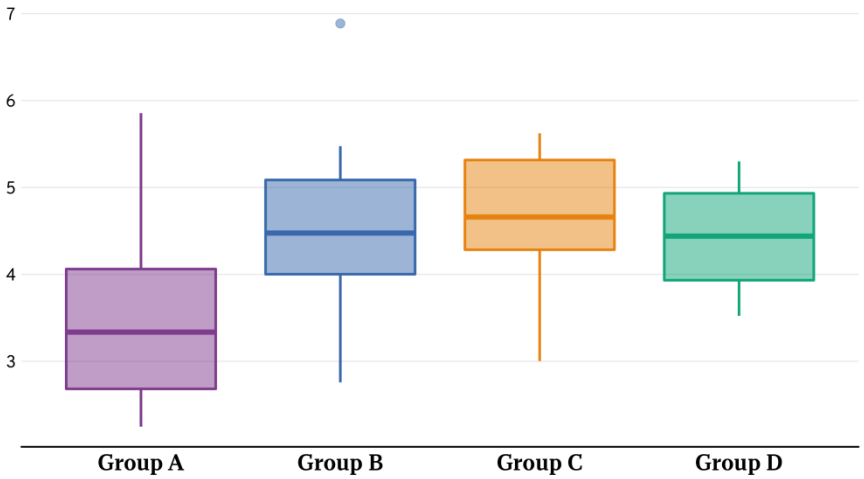
geom_boxplot()



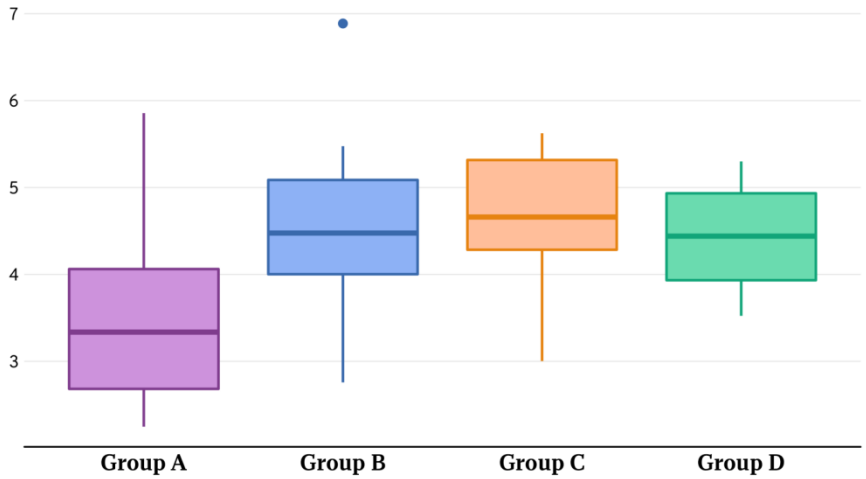
BOX-AND-WHISKERS PLOT

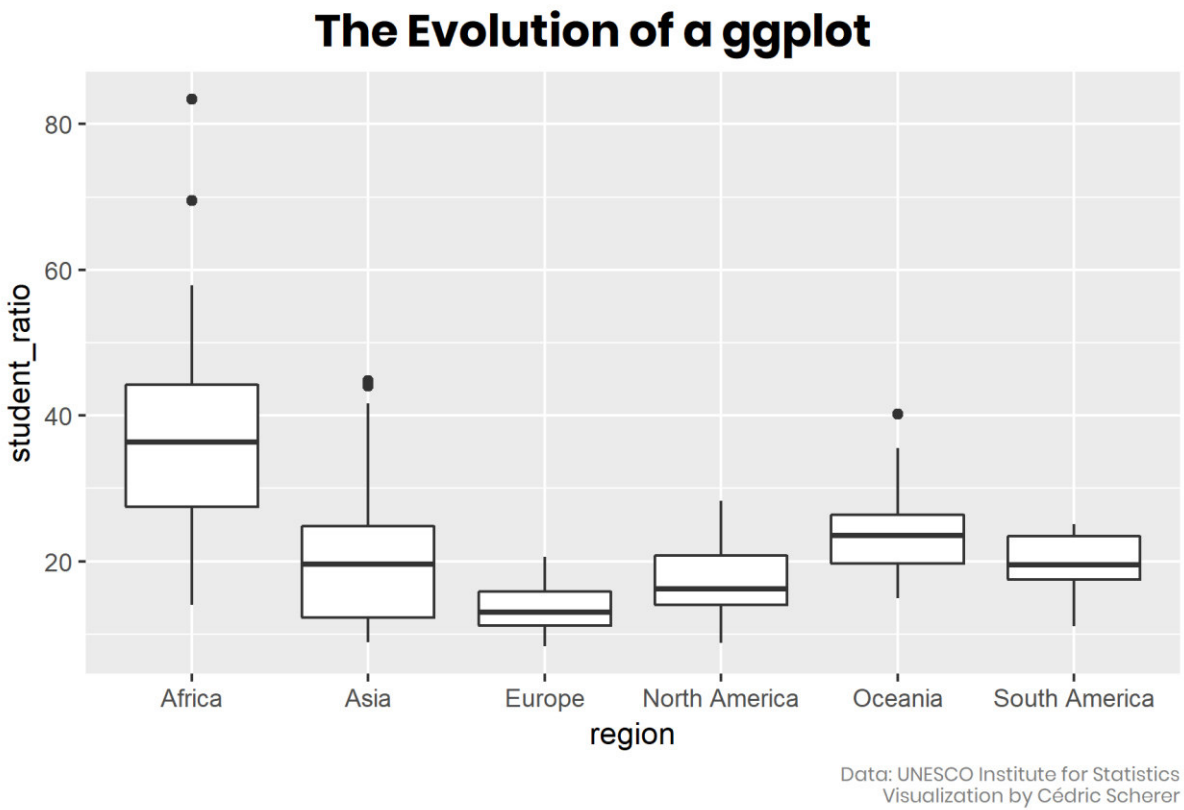
geom_boxplot()

```
aes(fill = group),
alpha = .5
```




```
aes(fill = group,
    fill = after_scale(colorspace::lighten(fill, .5)))
```

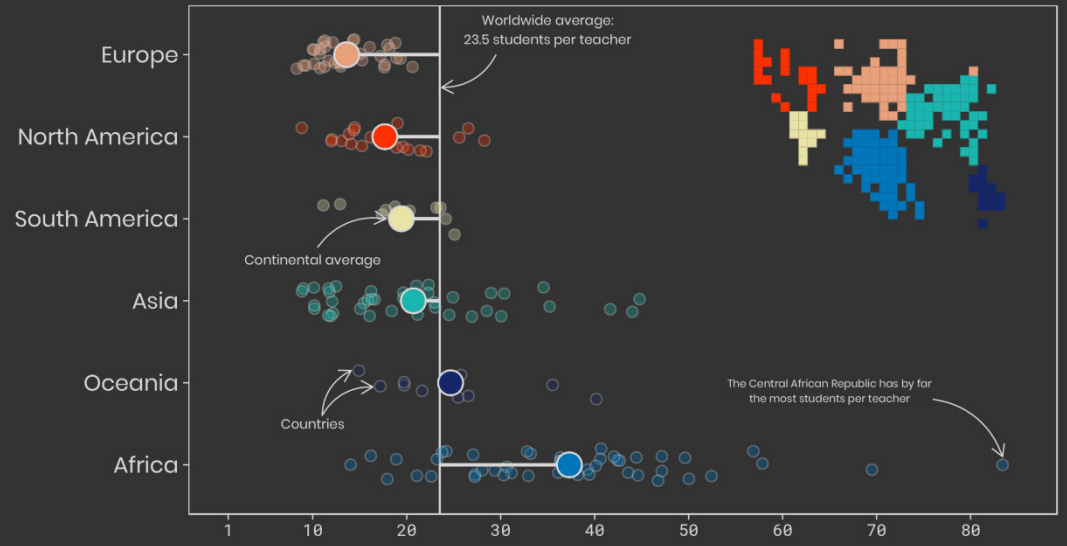


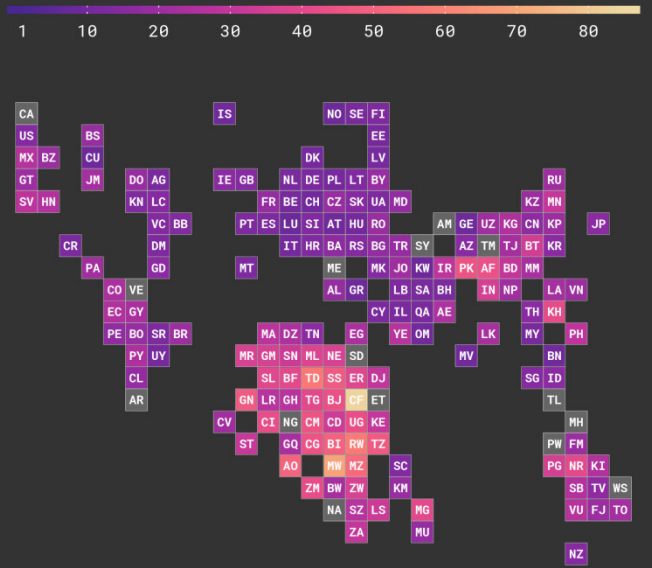


"The Evolution of a ggplot (Ep. 1)"

Global student to teacher ratios in primary education

Latest reported student to teacher ratio per country and continent (2012-2018)



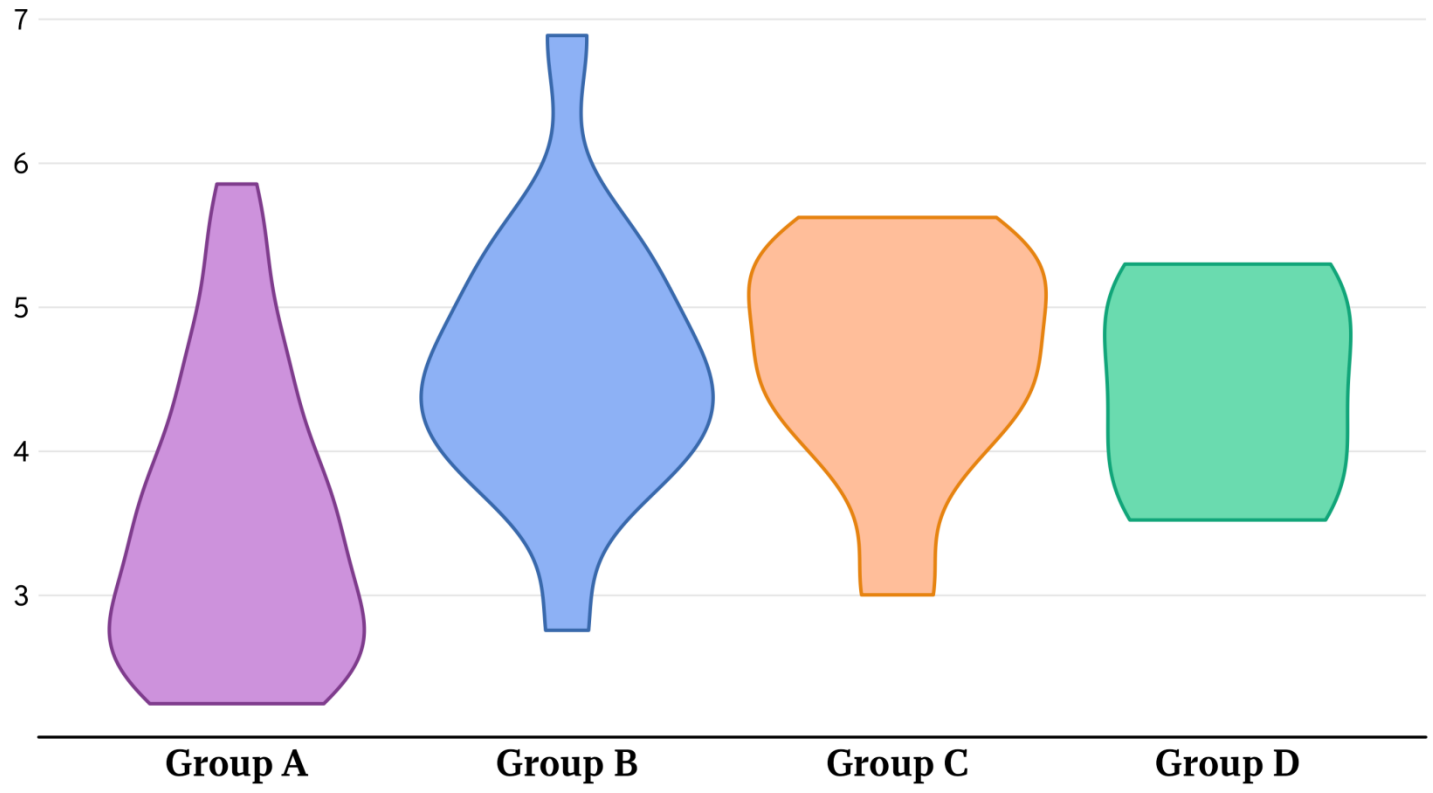


Visualization by Cédric Scherer | Data: "eAtlas of Teachers" by UNESCO

ALTERNATIVE CHART TYPES TO VISUALIZE *Distributions*

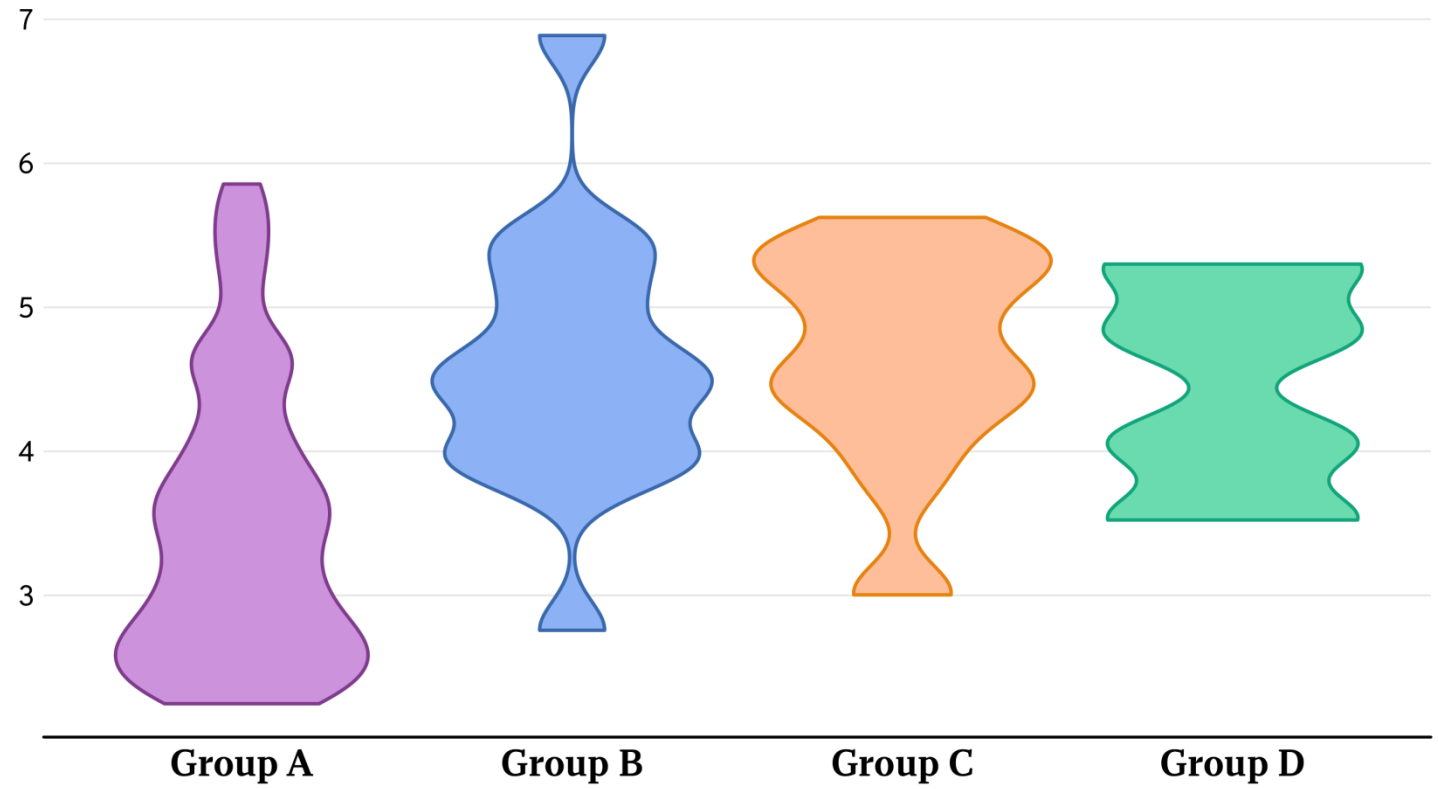
VIOLIN PLOTS

geom_violin()



VIOLIN PLOTS

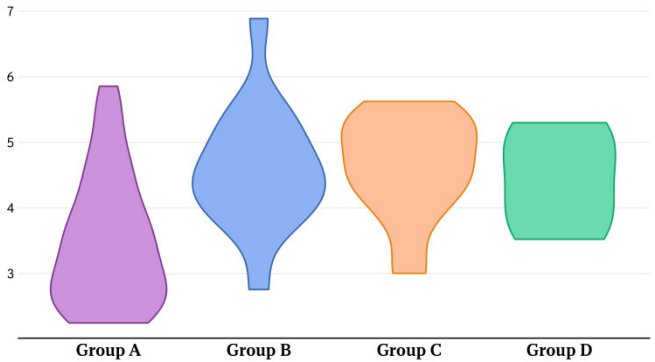
geom_violin(bw = .2)



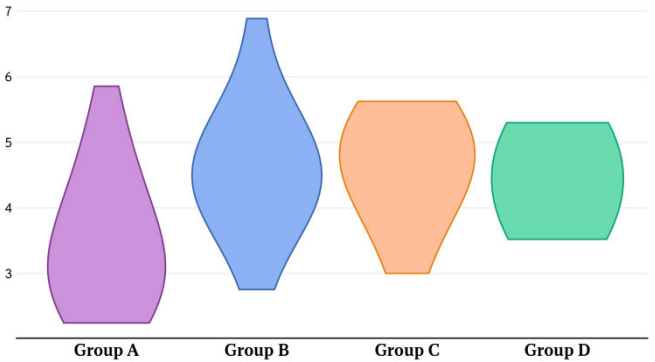
VIOLIN PLOTS

geom_violin()

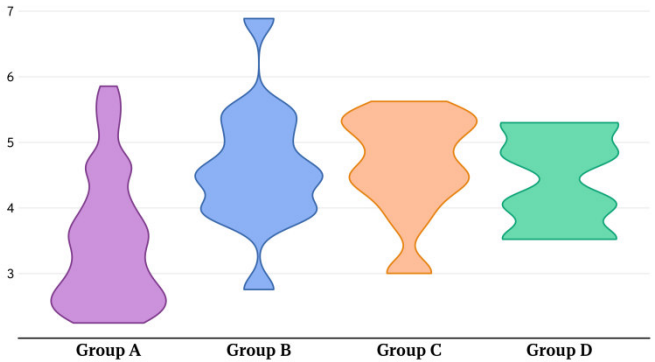
default



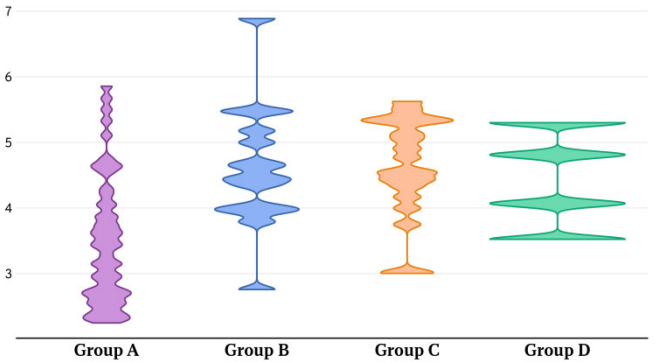
bw = 1



bw = .2

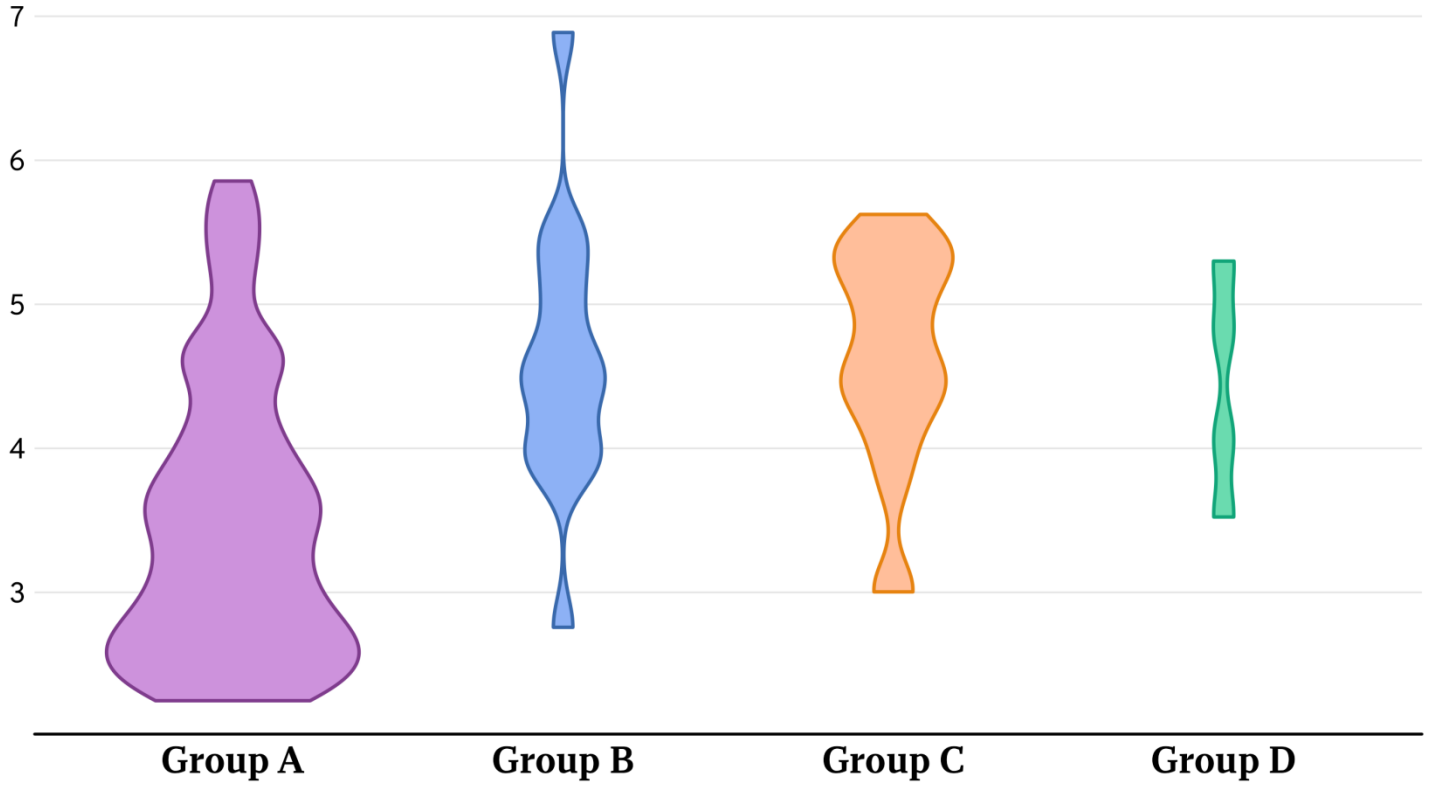


bw = .05

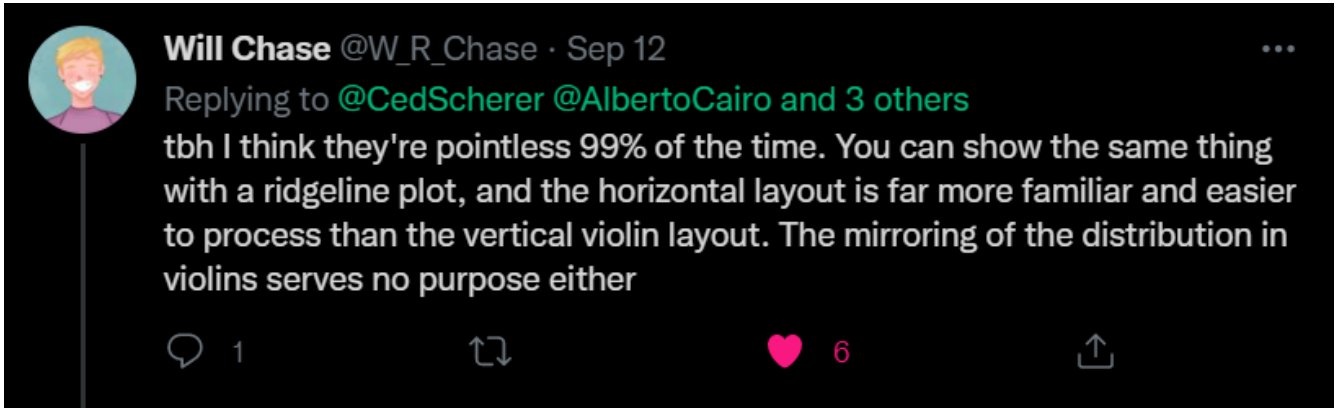


VIOLIN PLOT (SCALED BY COUNTS)

geom_violin(scale = "count")

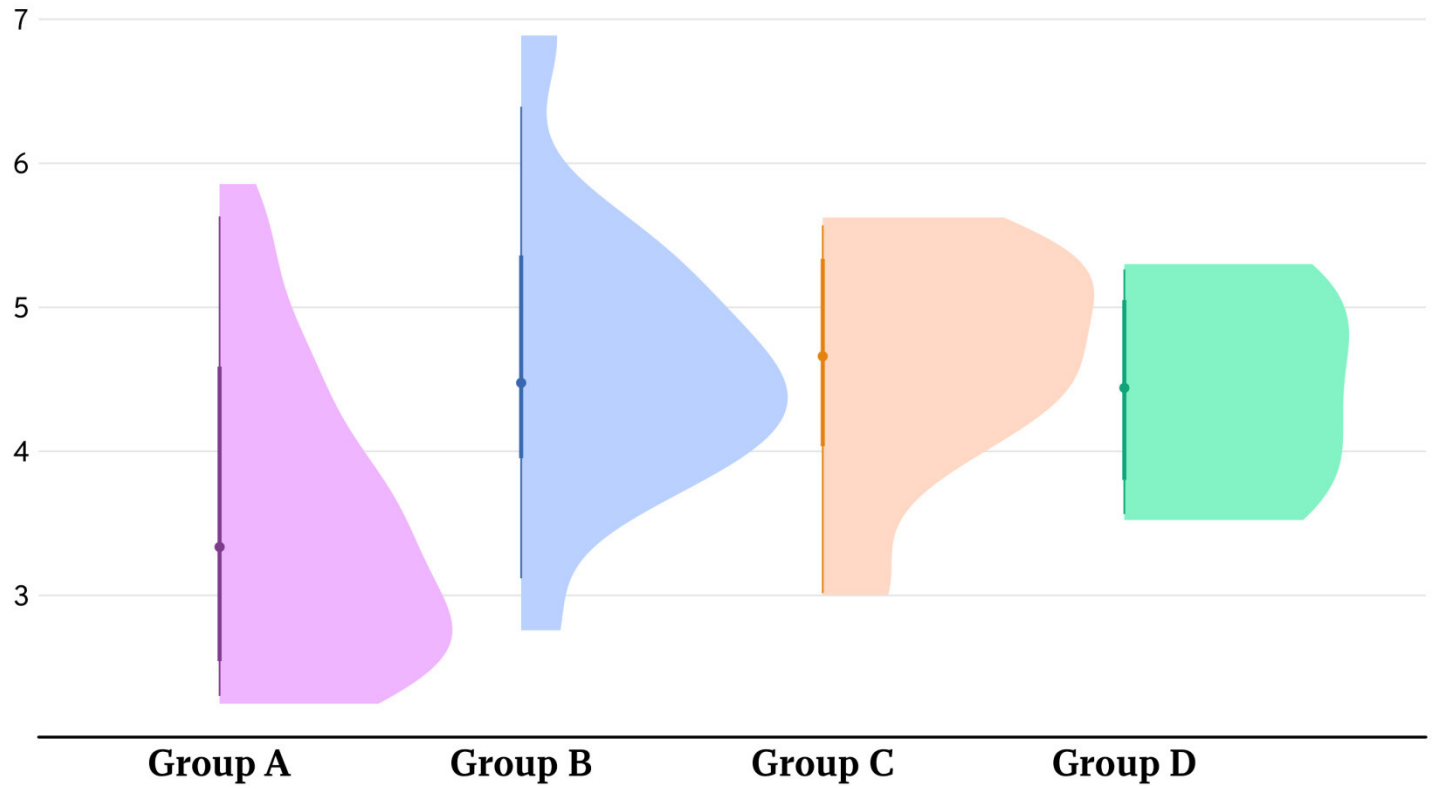


VIOLIN PLOTS IN THE WILD?



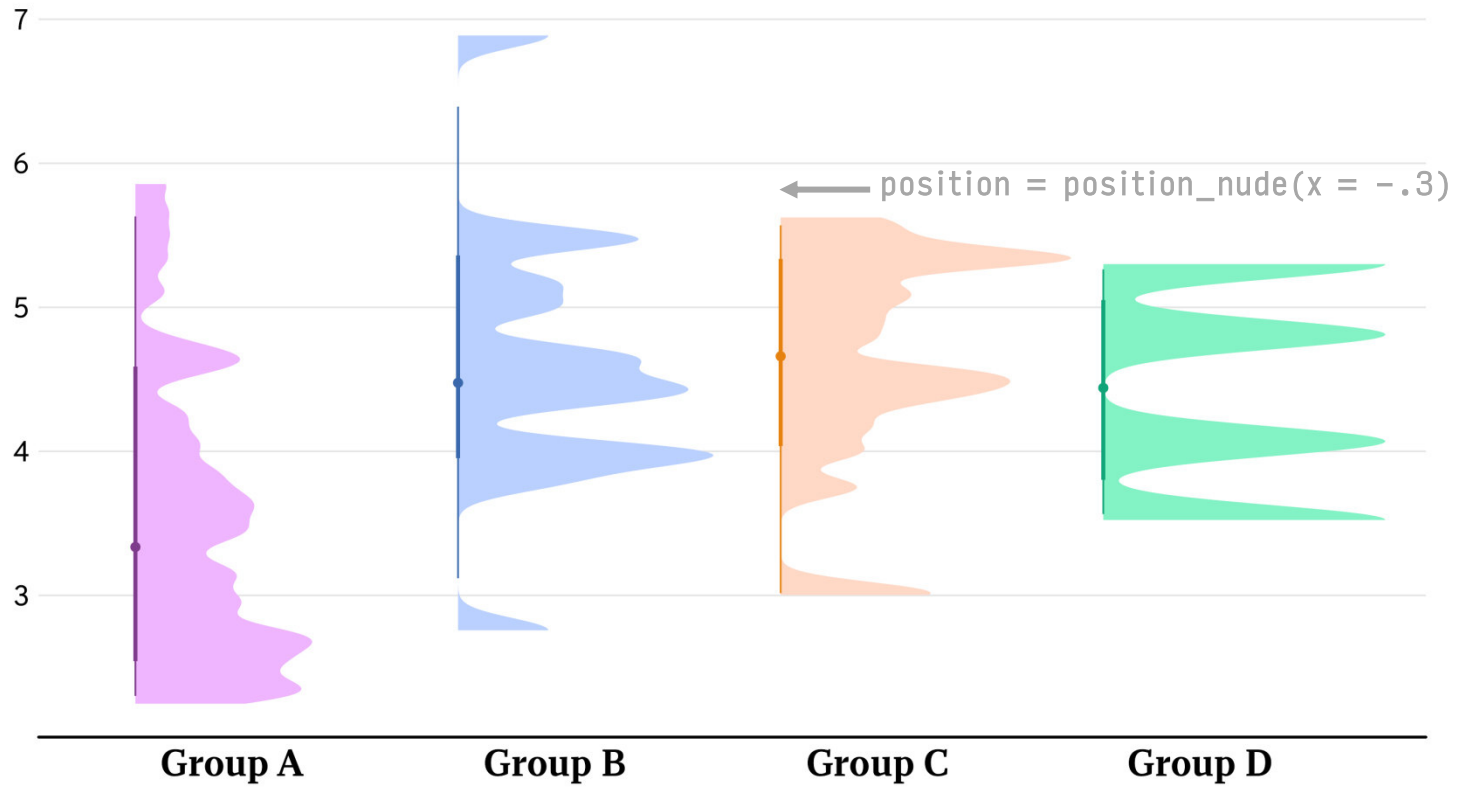
VIOLIN PLOT (HALF-EYES)

ggdist::stat_halfeye()



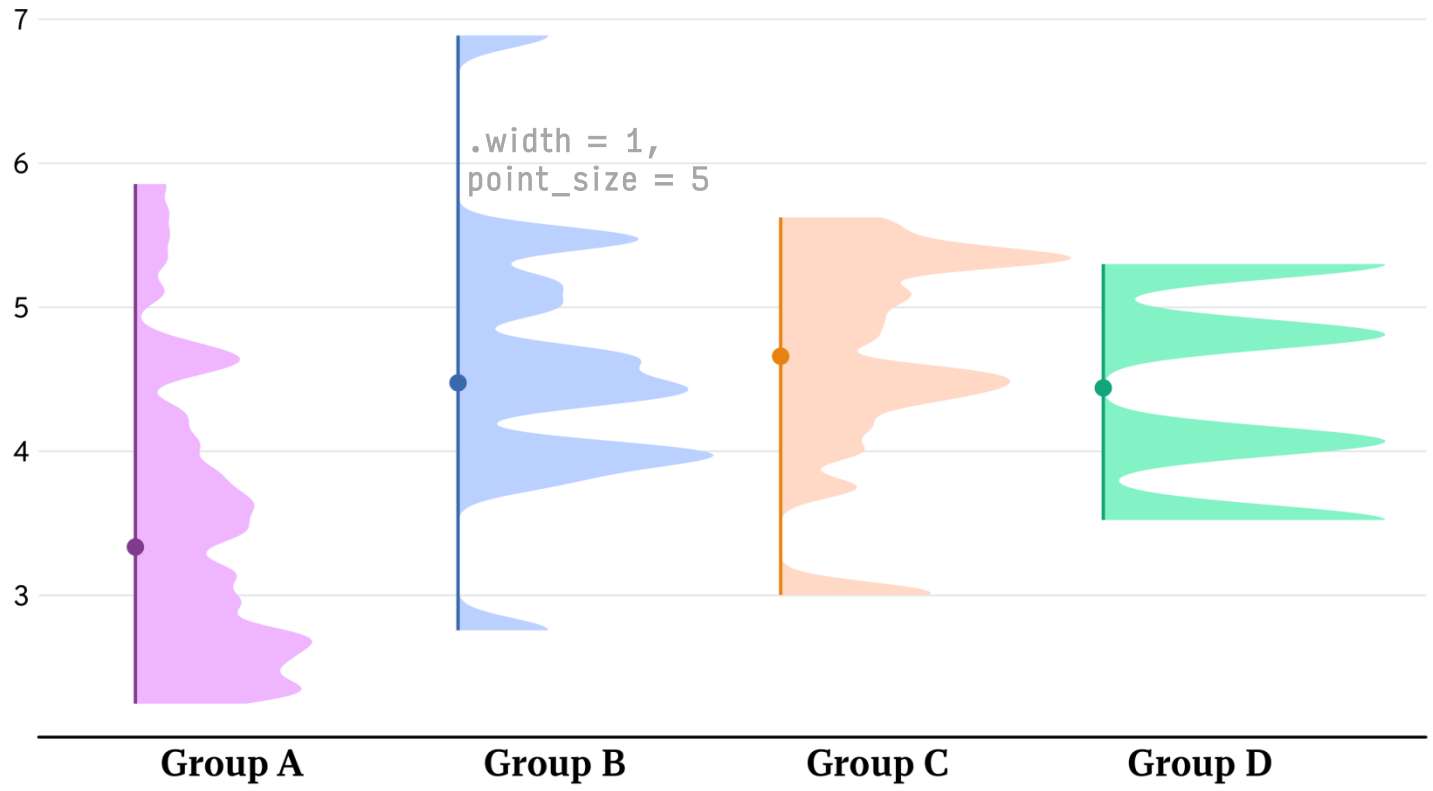
VIOLIN PLOT (HALF-EYES)

ggdist::stat_halfeye(adjust = .2)



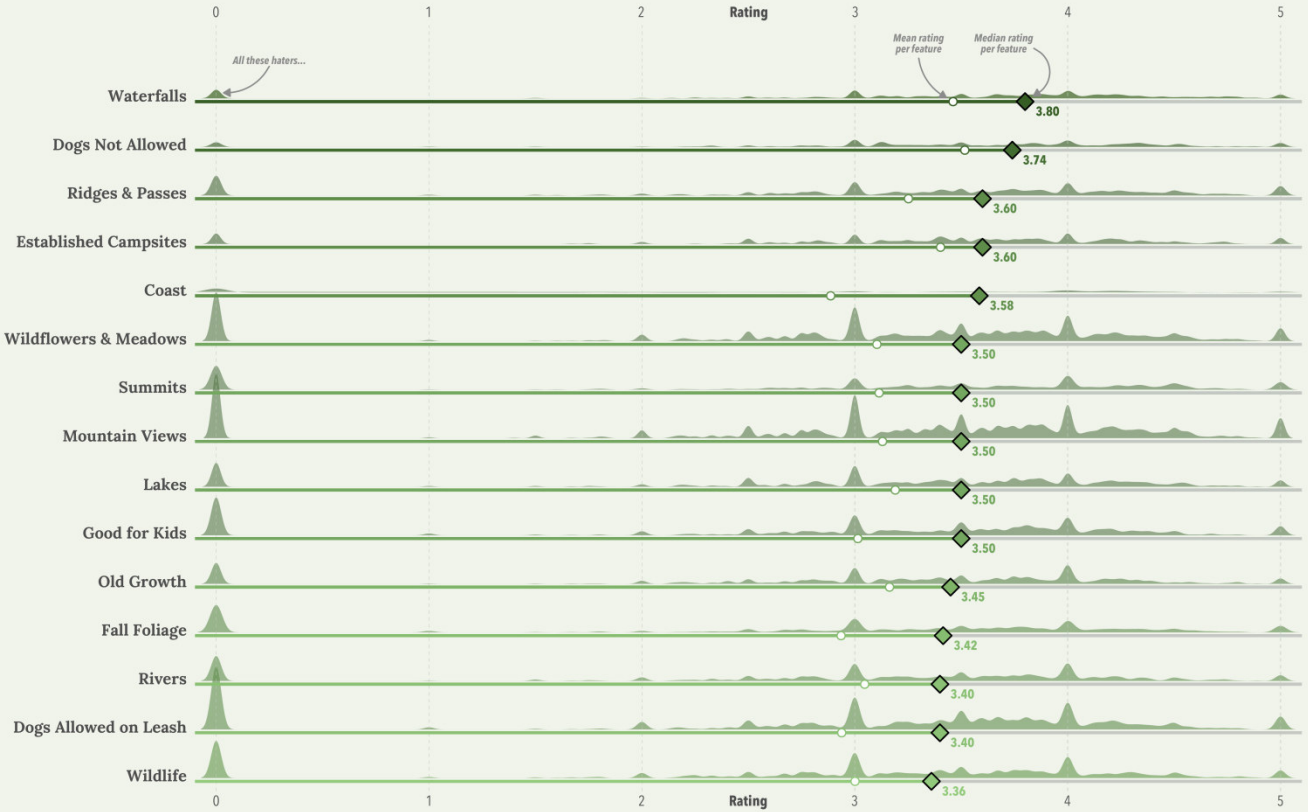
VIOLIN PLOT (HALF-EYES)

ggdist::stat_halfeye(adjust = .2)

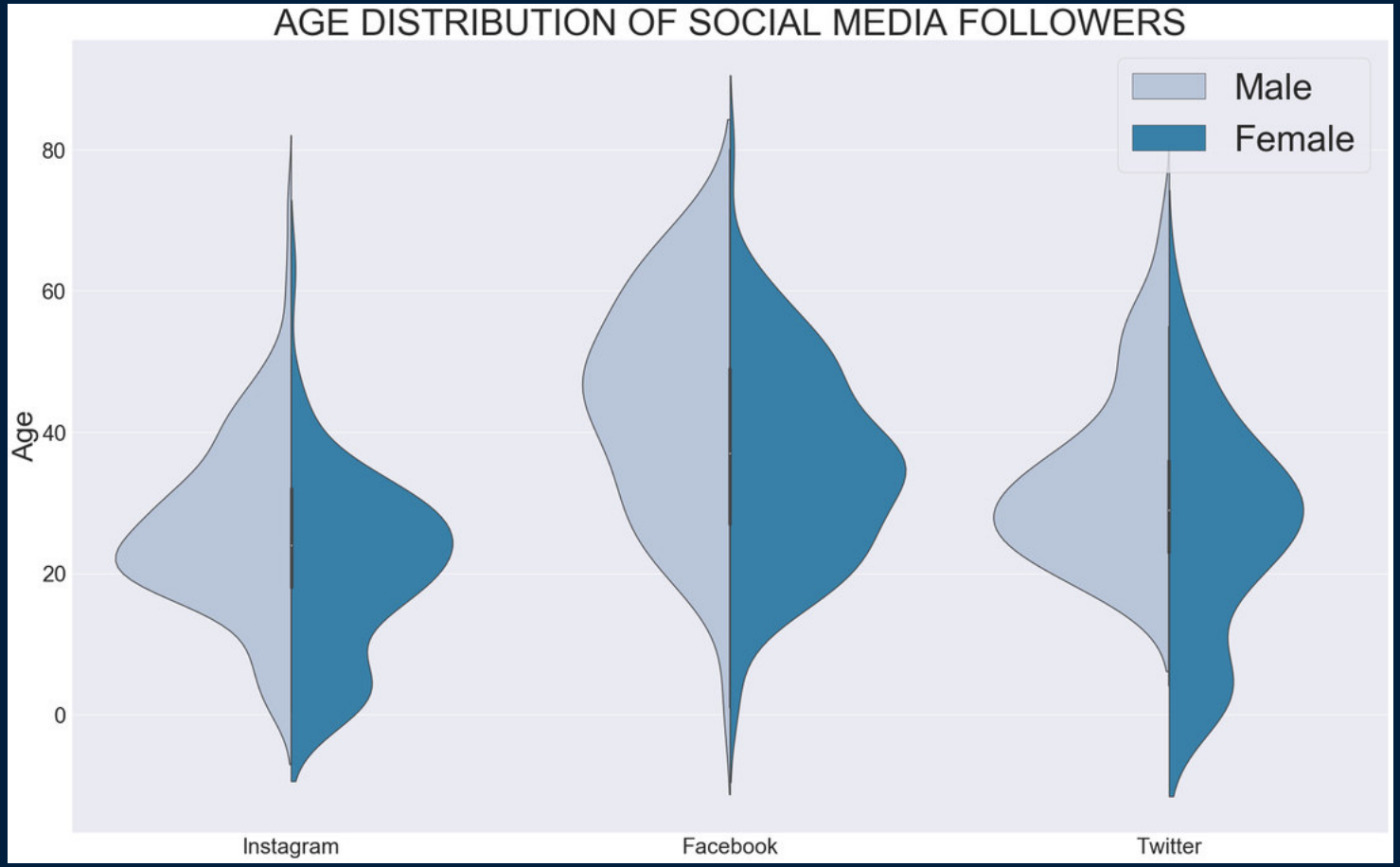


Hiking in Washington: Waterfalls are great—Wildlife and dogs rather not

The **Washington Trait Association (WTA)** offers a hiking guide that claims top be *"the most comprehensive database of hikes in Washington"*. It comprises content written by local hiking experts and user submitted information such as ratings for each trail and specification of certain features. The plot shows the overall rating of 1,924 Washington trails listed in the WTA database grouped by its features and ordered by the highest median rating.



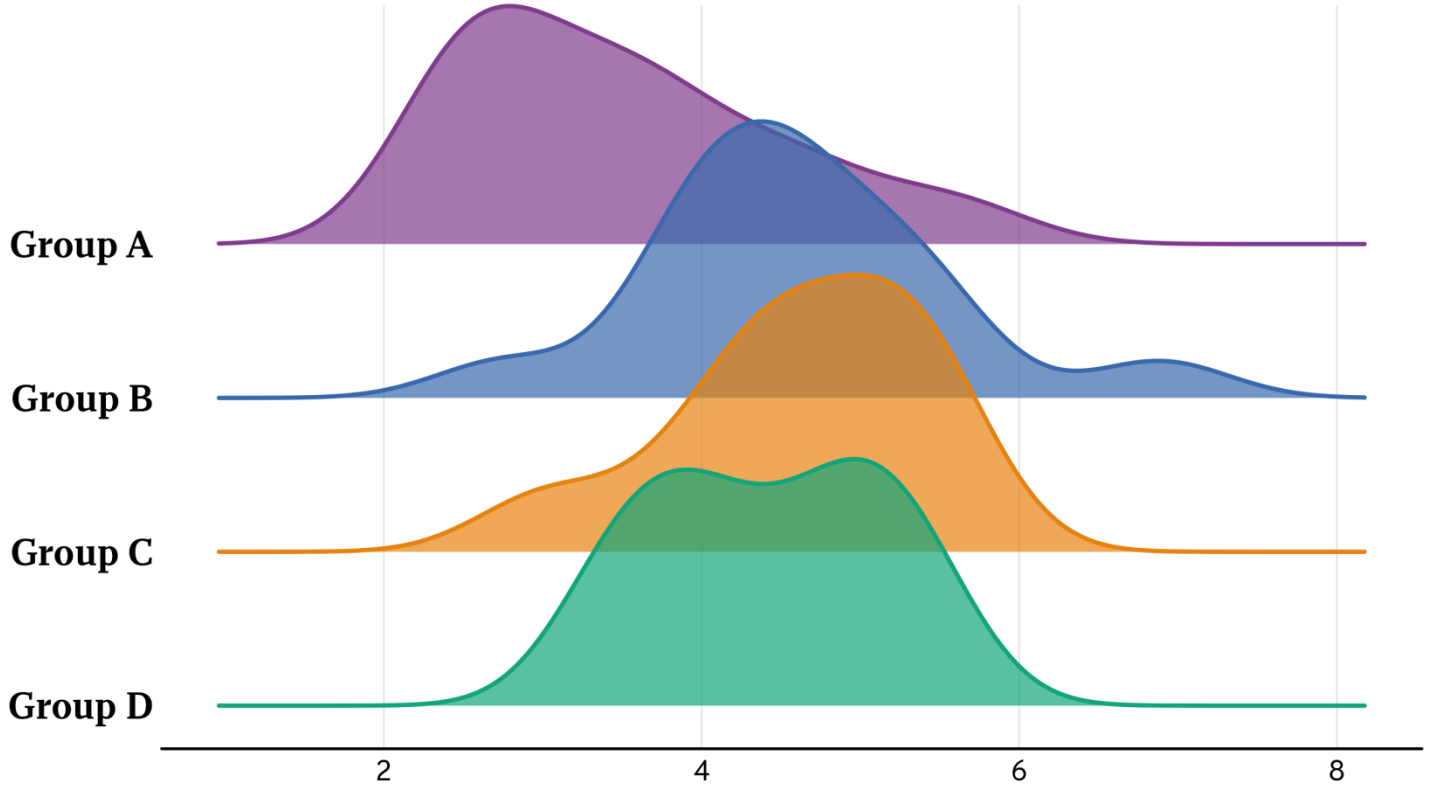
Visualization by Cédric Scherer • Data by WTA



"Violin Plots" by LondonSoda

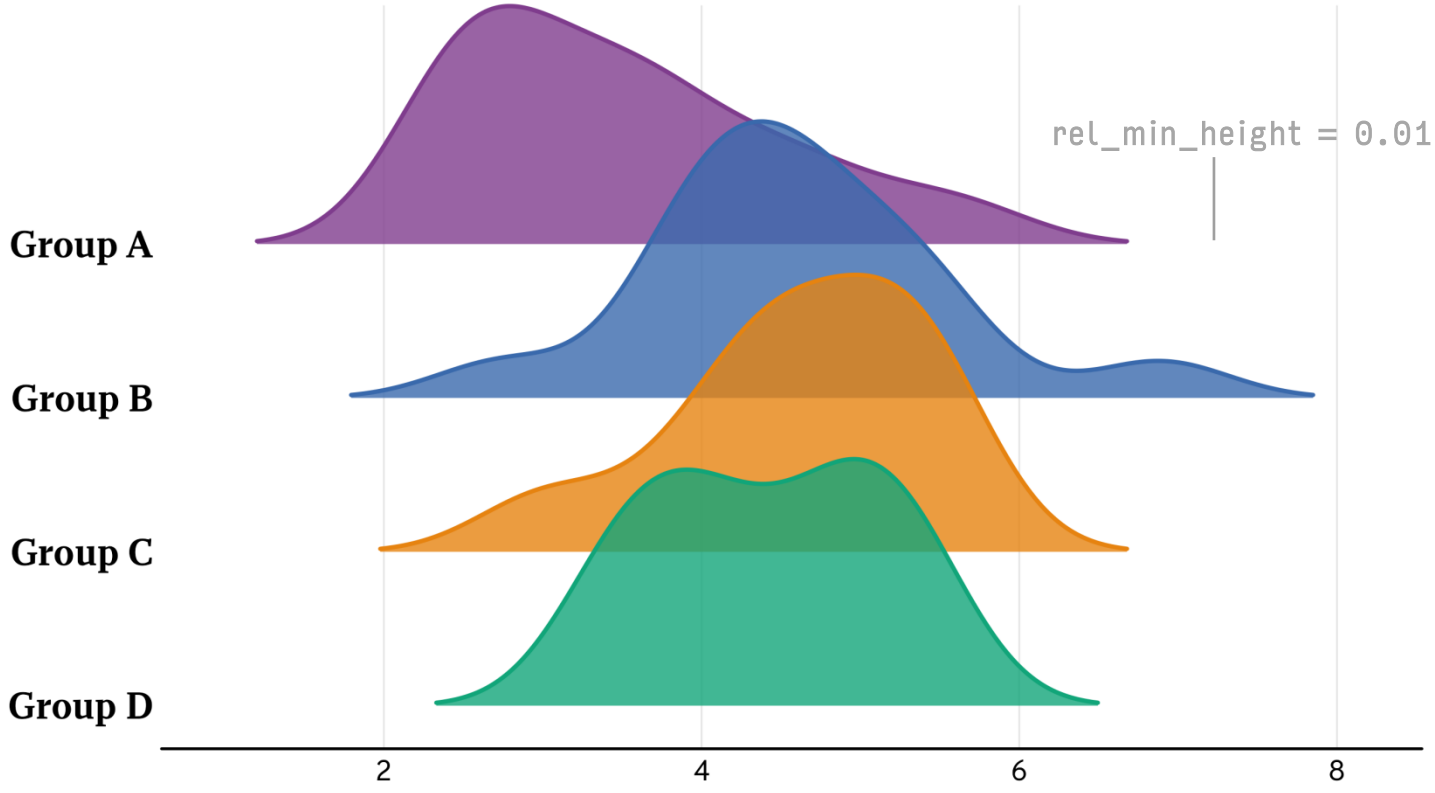
RIDGELINE PLOT

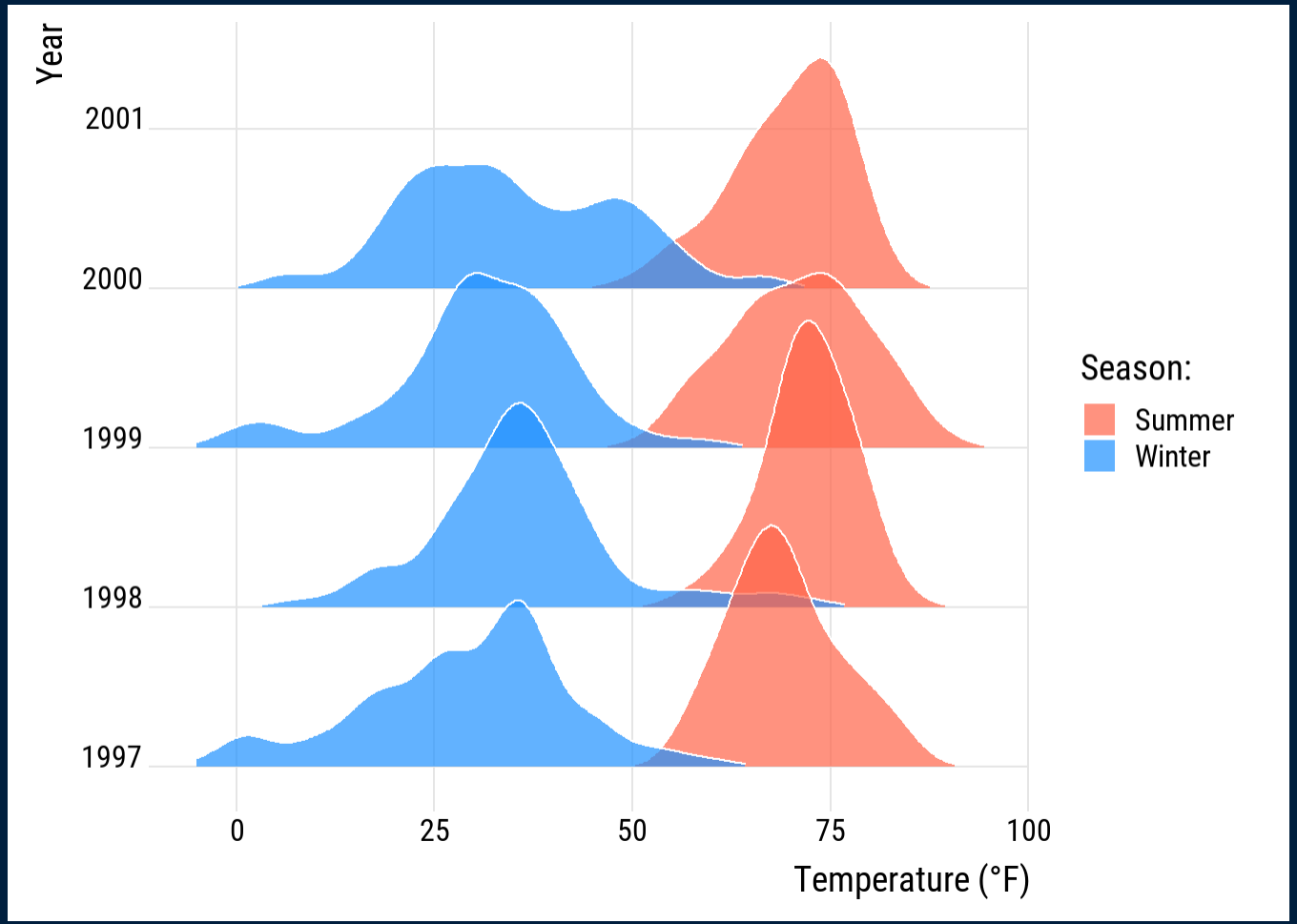
ggridges::geom_density_ridges()



RIDGELINE PLOT (CUTTED)

ggridges::geom_density_ridges()

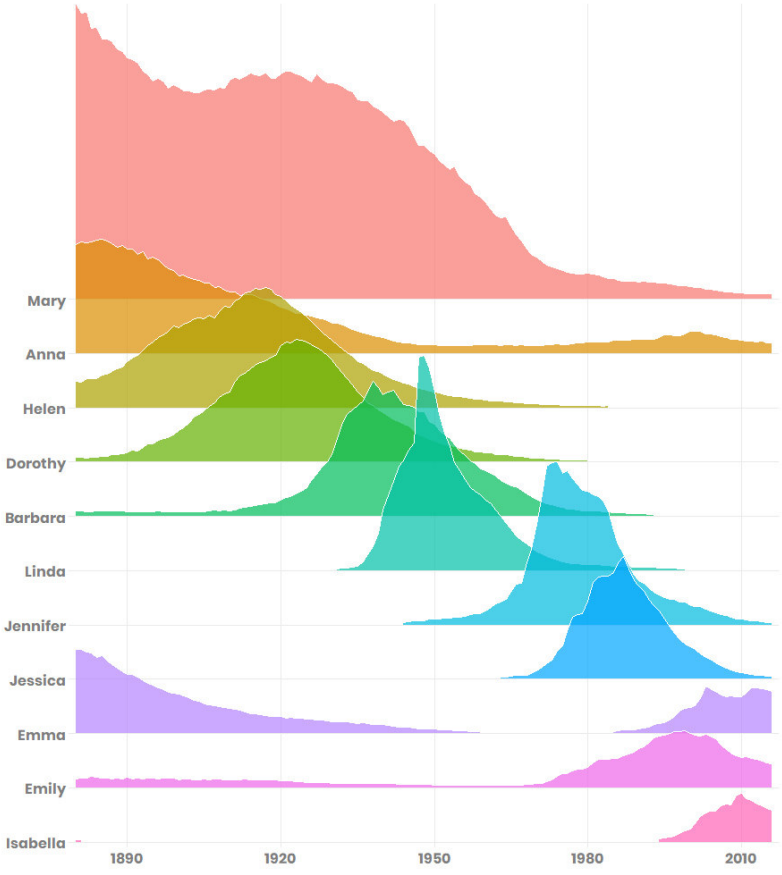




"A ggplot2 Tutorial for Beautiful Plotting in R"

Most popular girl names in the U.S.

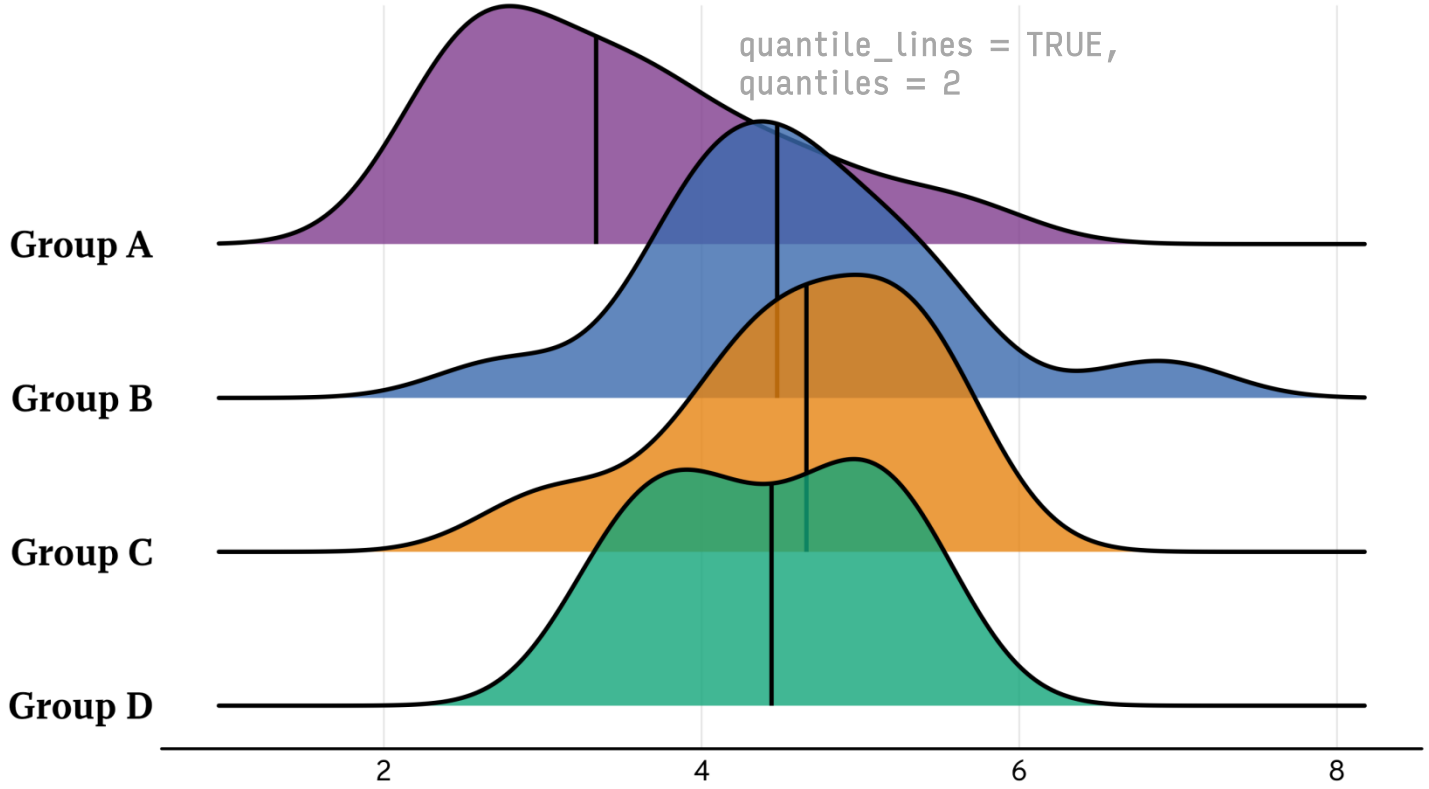
Top 2 names with the highest mean and/or maximum per quarter are shown.



Source: U.S. Social Security Administration

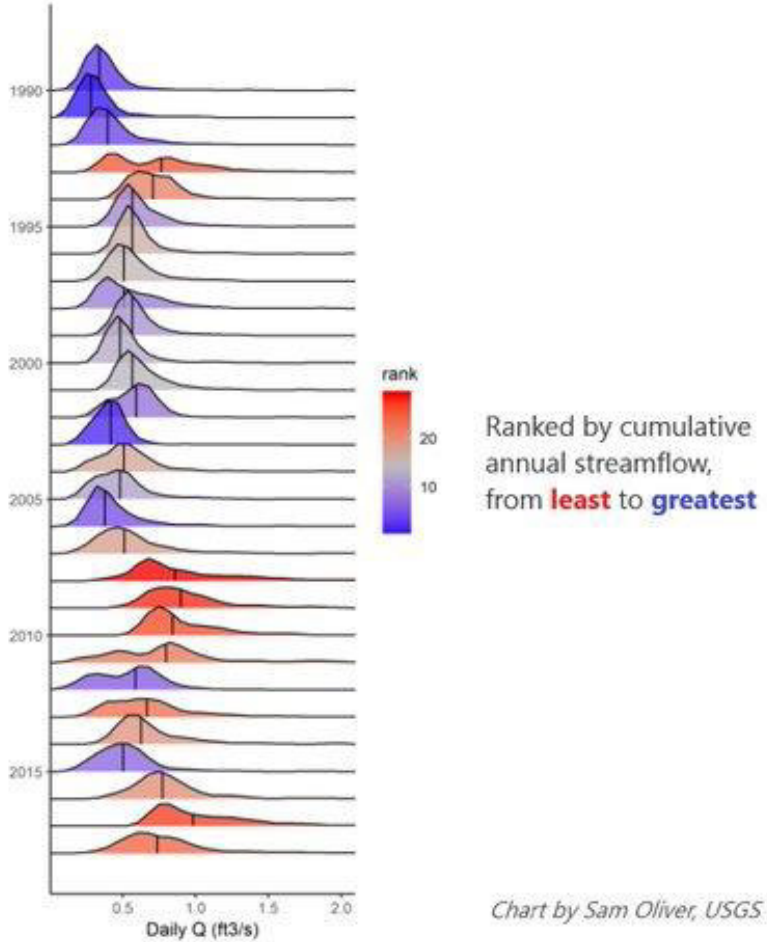
RIDGELINE PLOT WITH MEDIAN

ggridges::stat_density_ridges()



Distributions of daily streamflow on the Yahara River by year

1990 - 2018



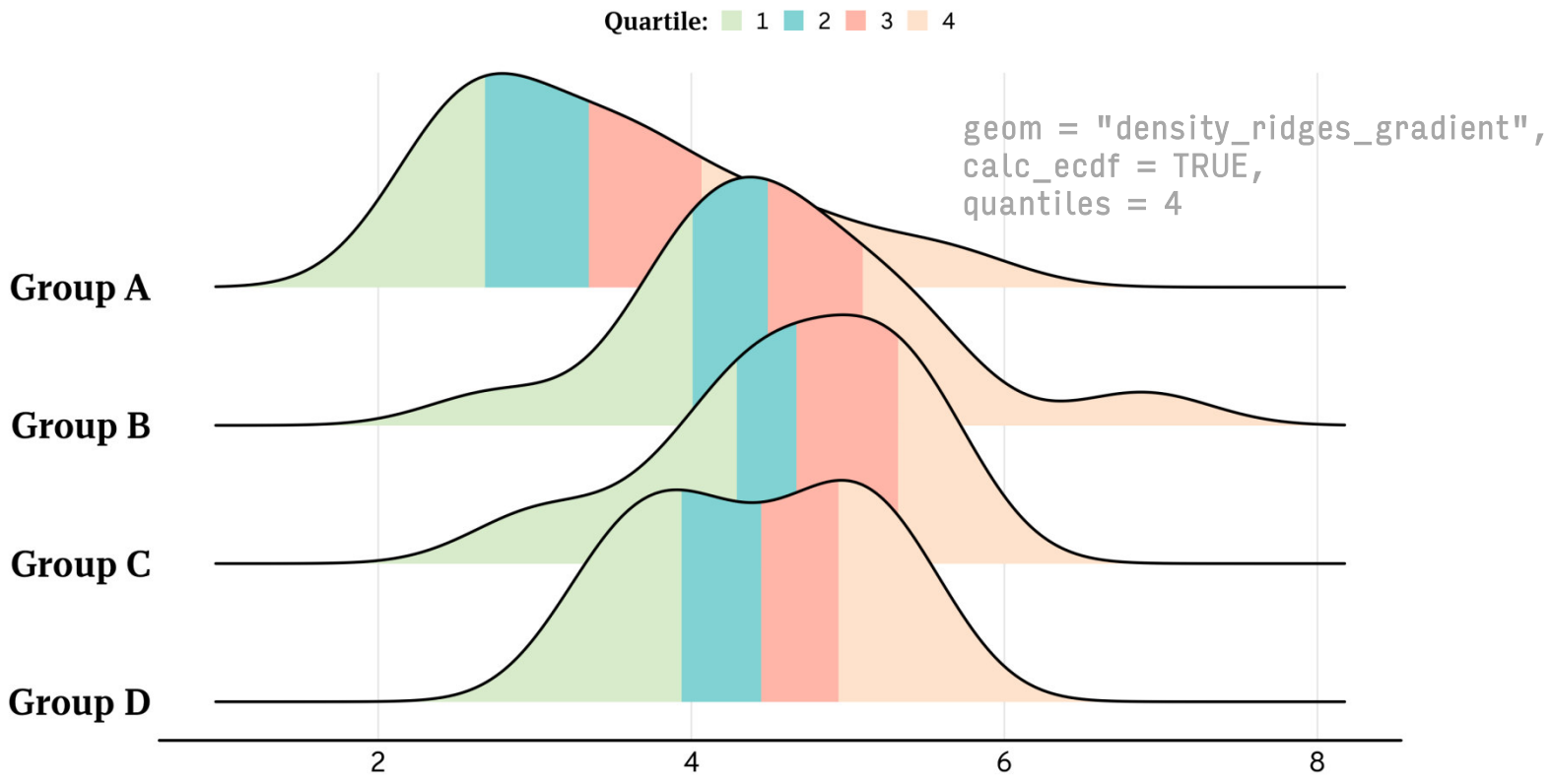
Ranked by cumulative annual streamflow, from **least** to **greatest**

Chart by Sam Oliver, USGS

Contribution to #30DayChartChallenge by @USGS_DataSci

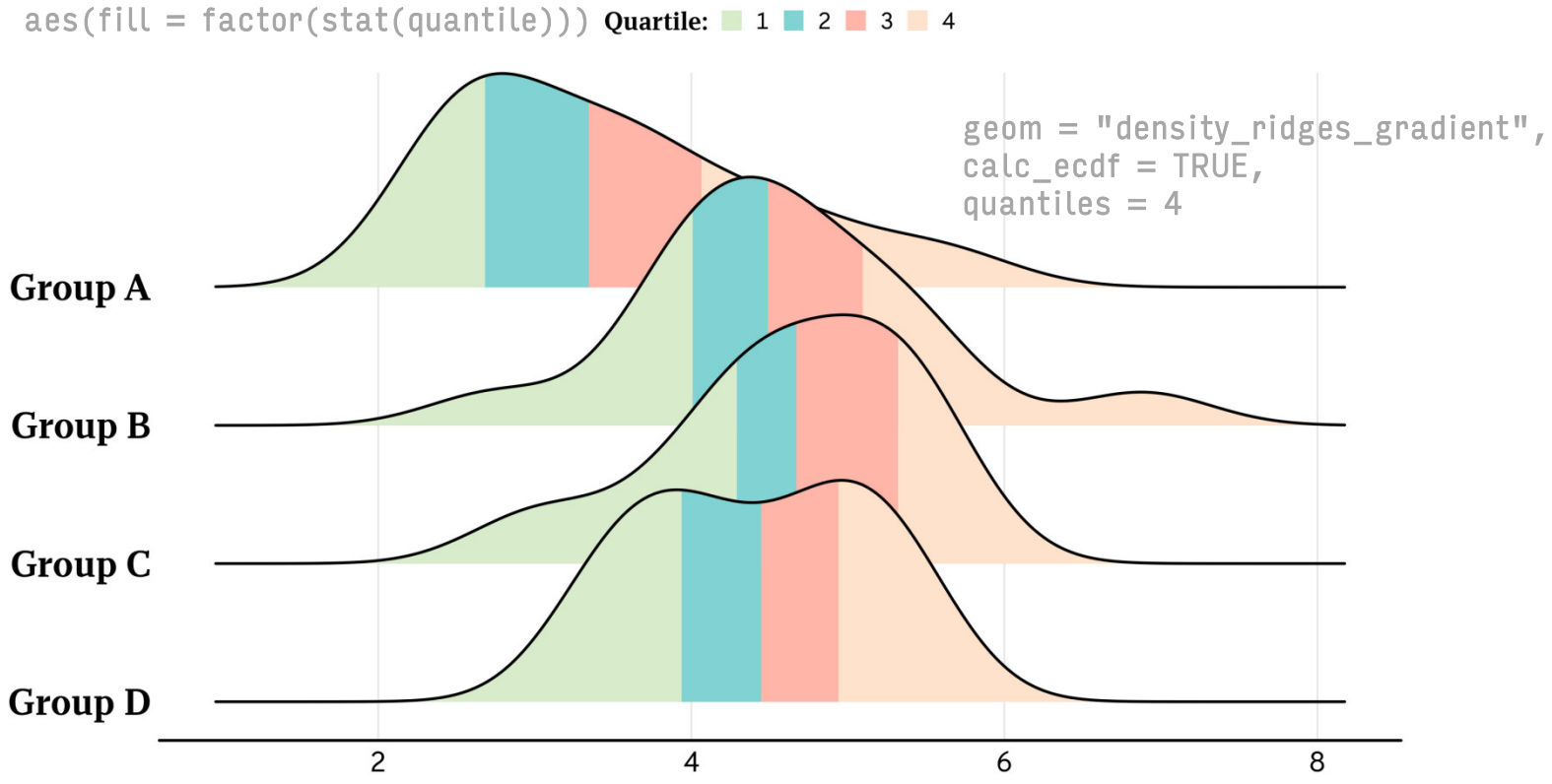
RIDGELINE PLOT WITH QUARTILES

ggridges::stat_density_ridges()



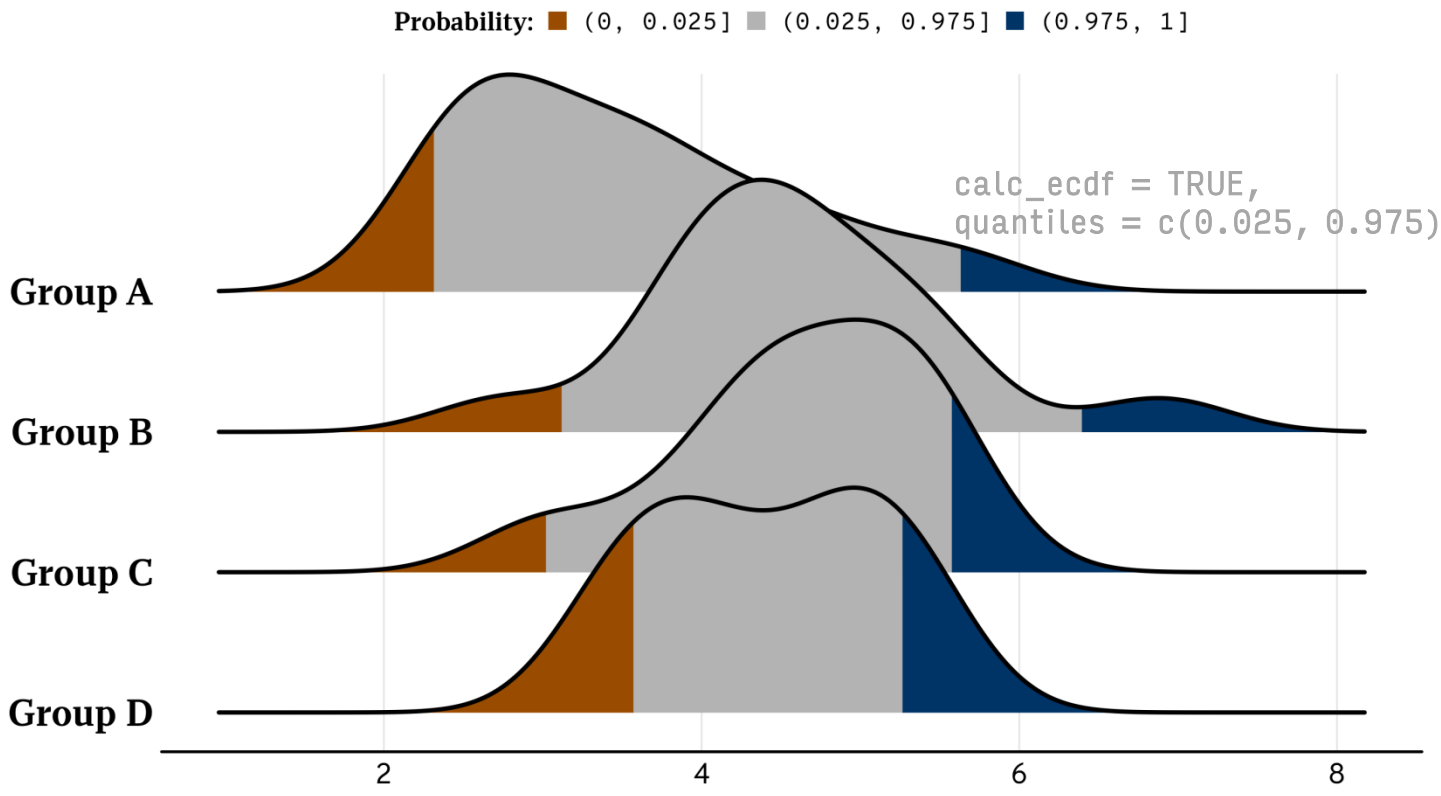
RIDGELINE PLOT WITH QUARTILES

ggridges::stat_density_ridges()



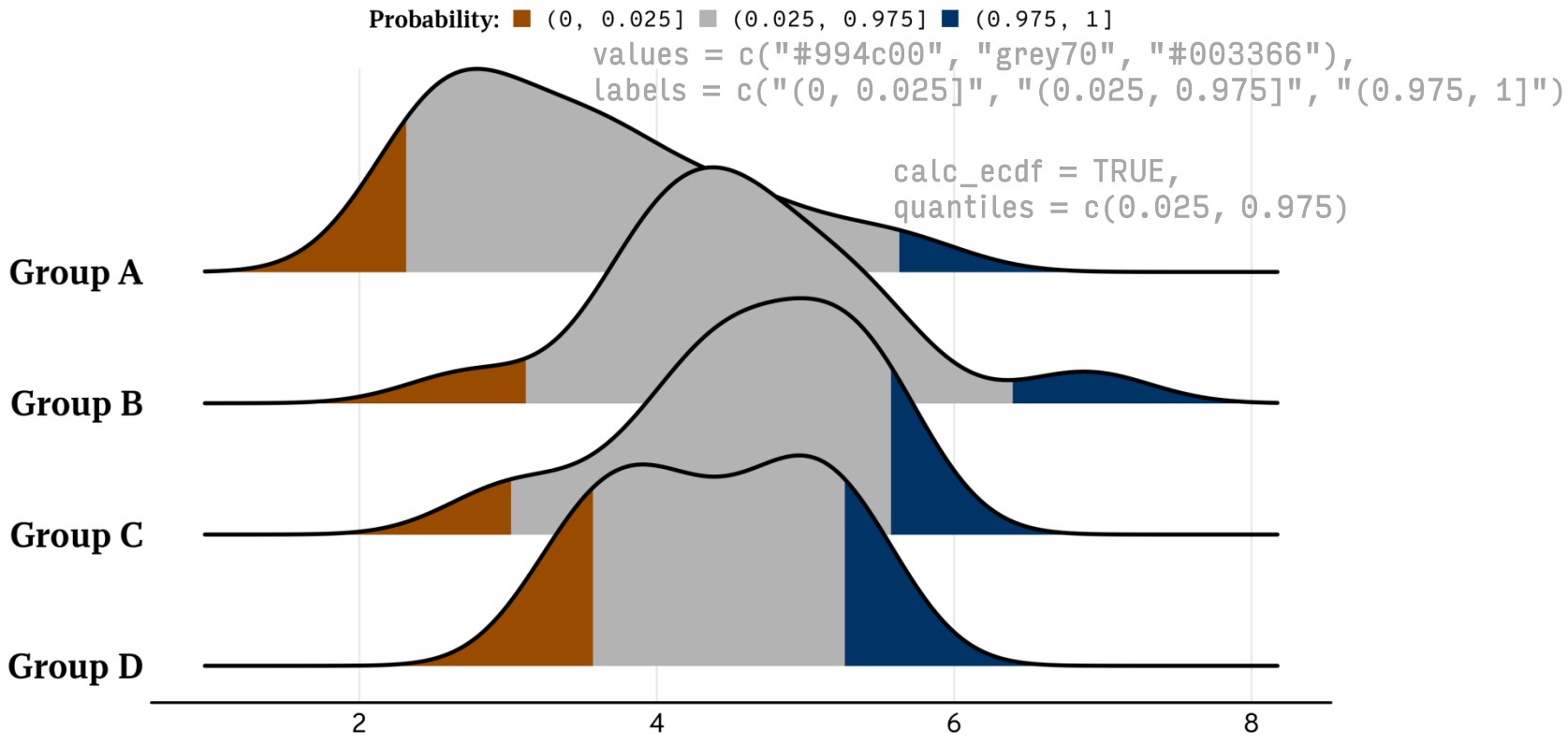
RIDGELINE PLOT WITH MEDIAN

ggridges::stat_density_ridges()



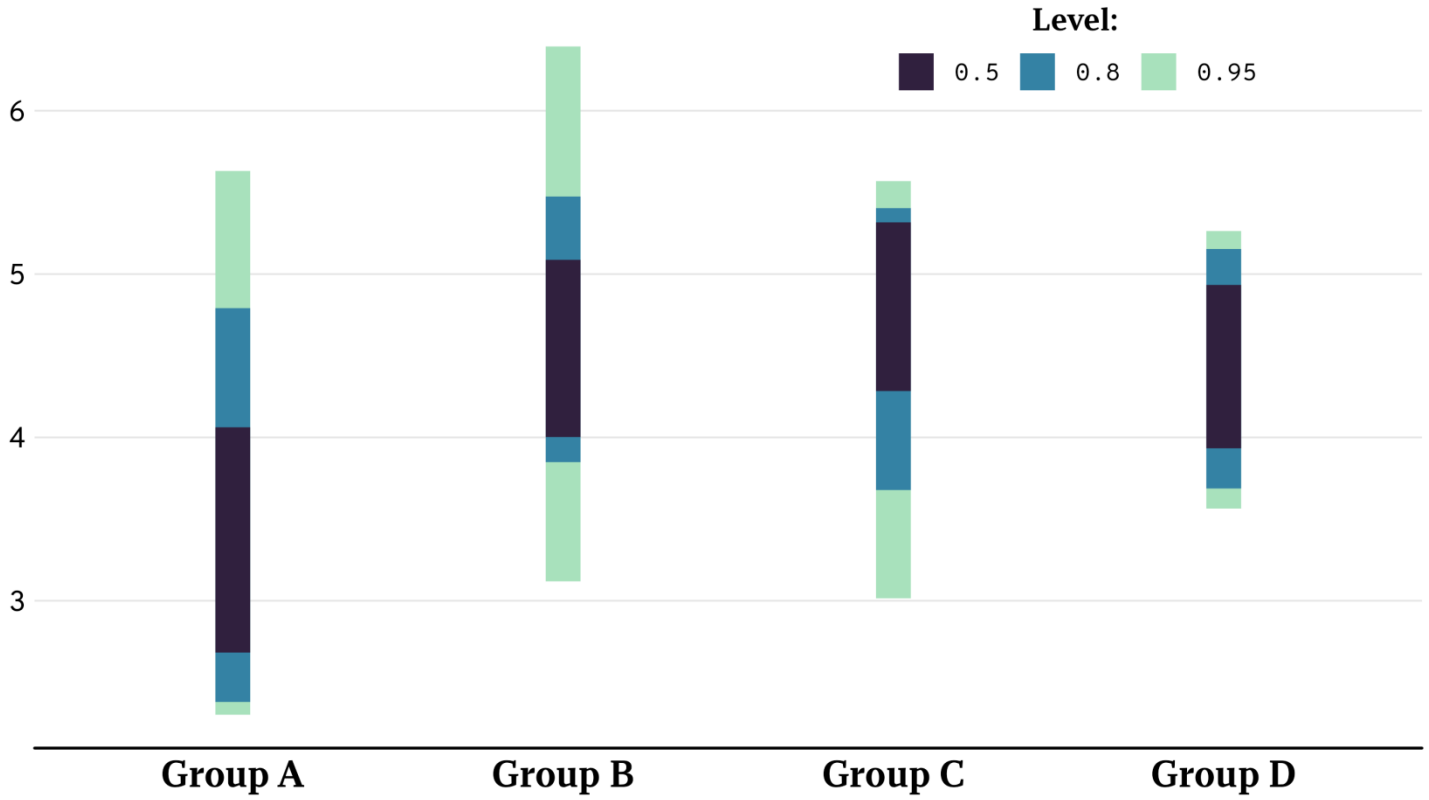
RIDGELINE PLOT WITH MEDIAN

ggridges::stat_density_ridges()



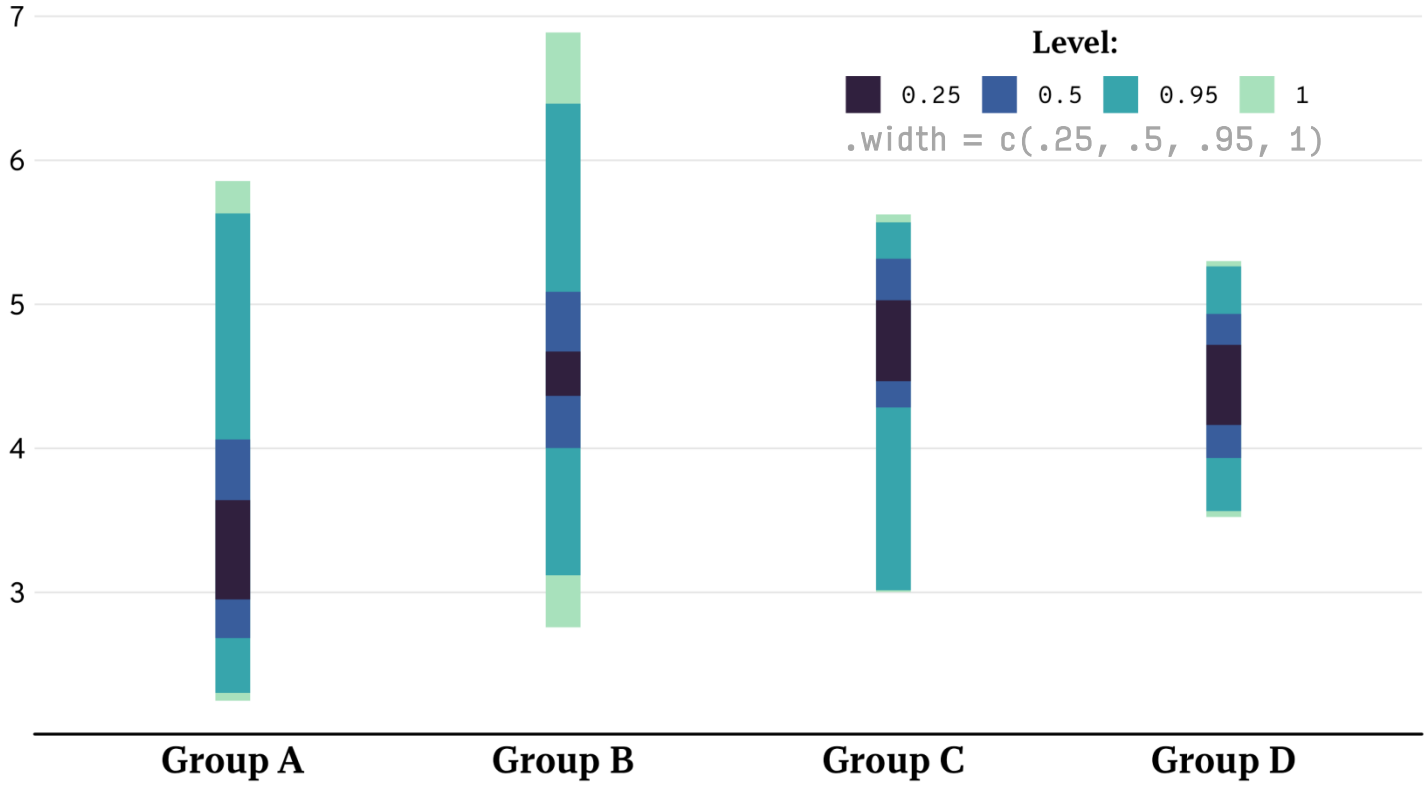
INTERVAL STRIPS

ggdist::stat_interval()



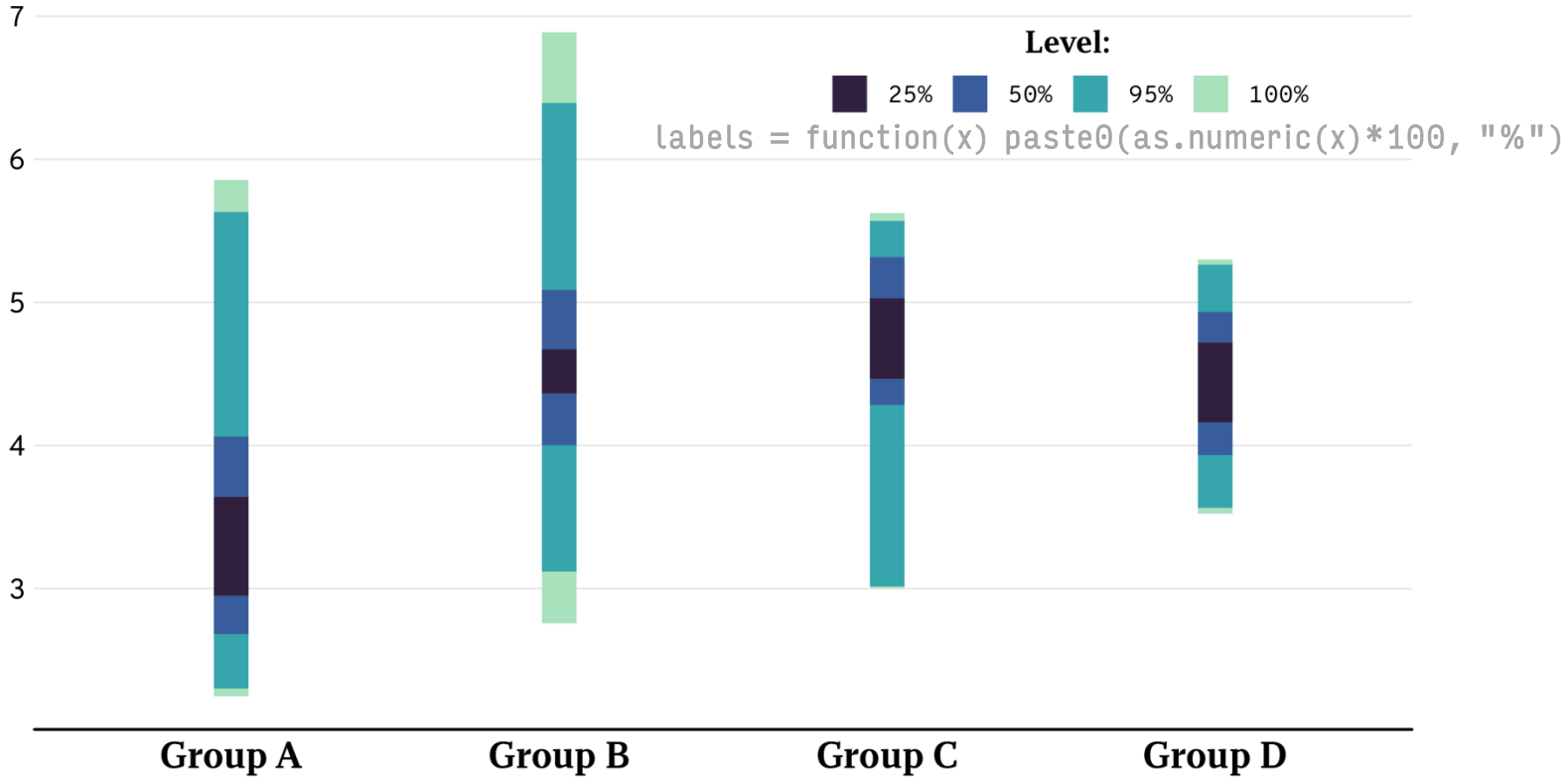
INTERVAL STRIPS

ggdist::stat_interval()



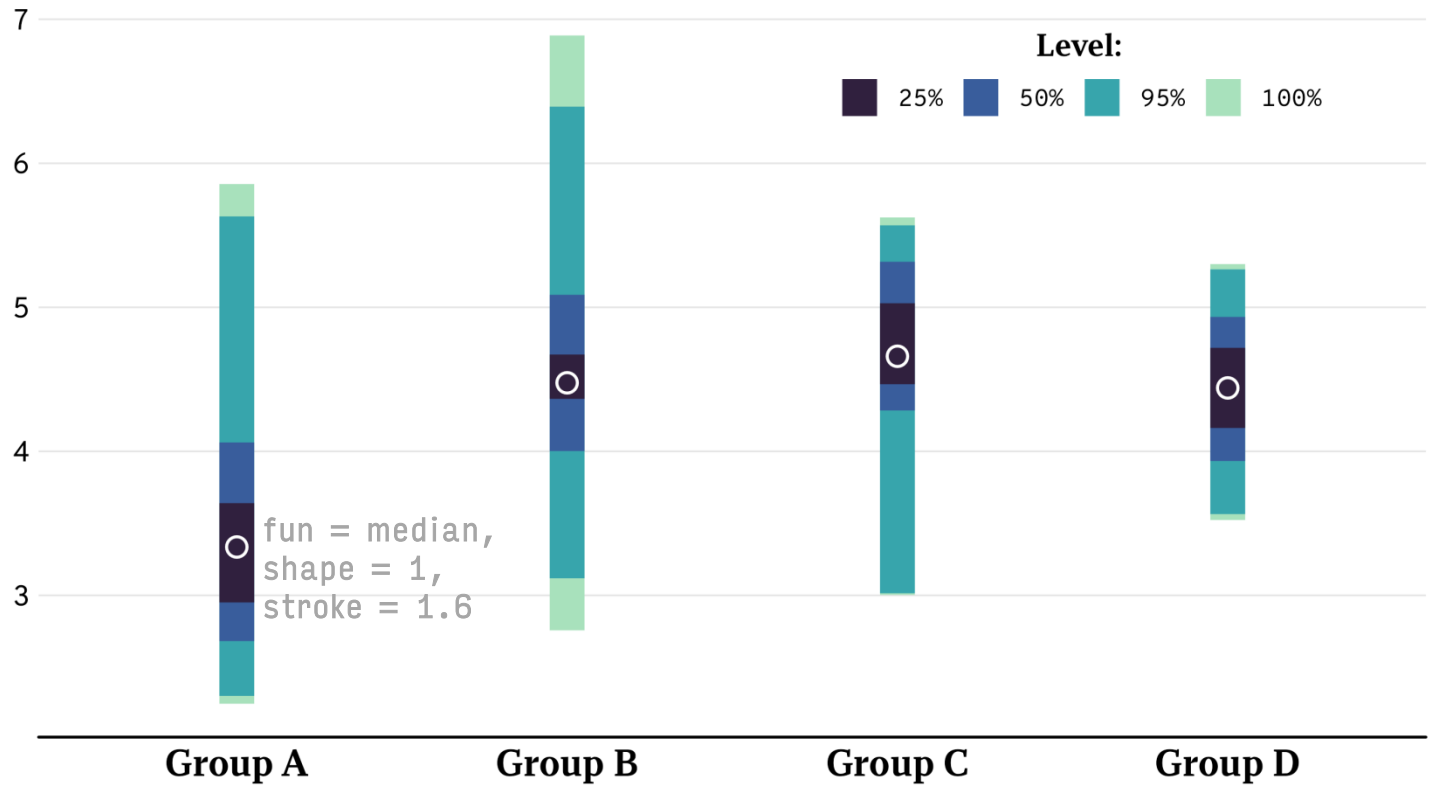
INTERVAL STRIPS

ggdist::stat_interval()



INTERVAL STRIPS

ggdist::stat_interval() + stat_summary(geom = "point")

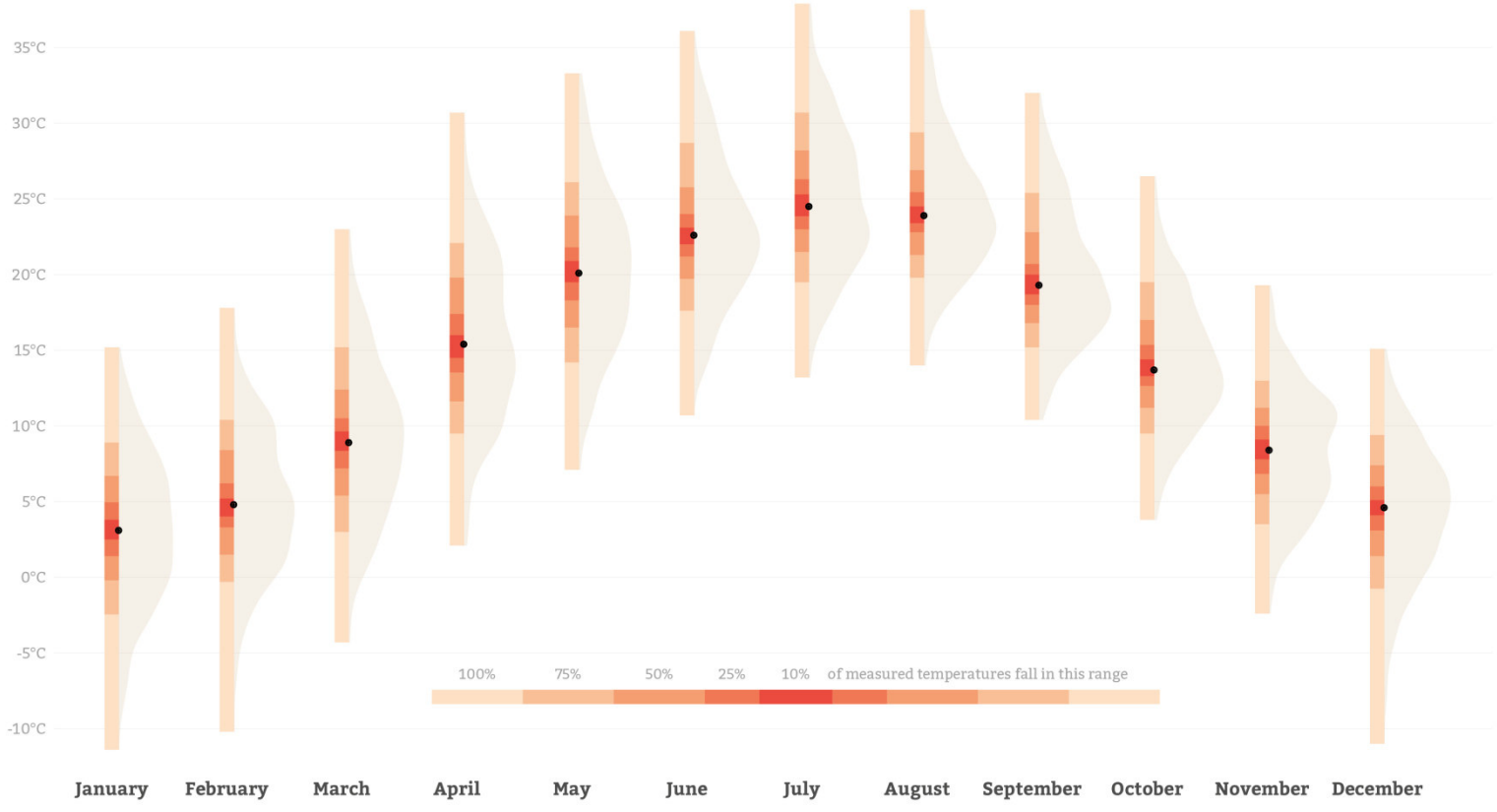


INTERVAL STRIPS X HALF-EYE

ggdist::stat_interval() + ggdist::stat_halfeye()

Daily Temperatures in Berlin, Germany

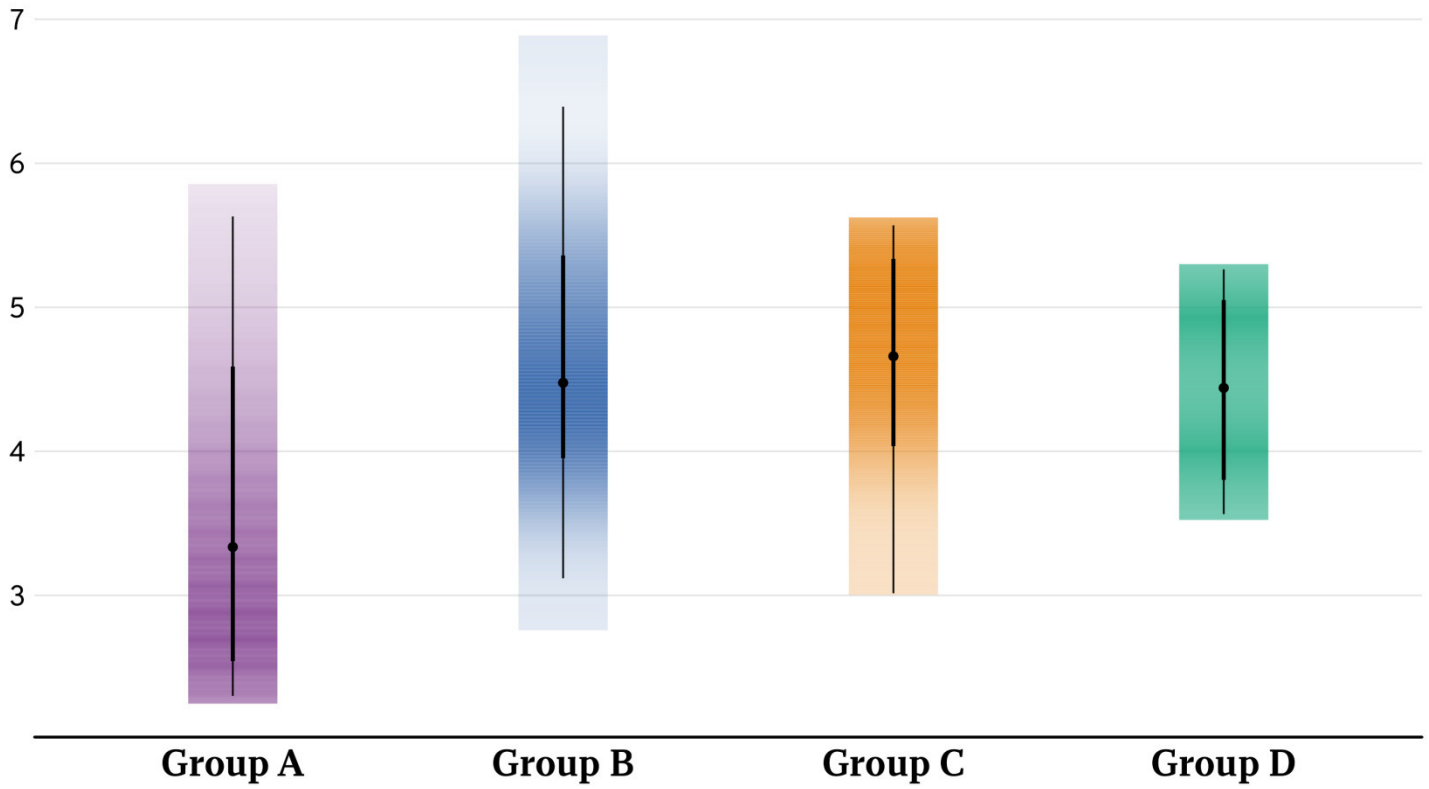
Range and distribution of maximum daily temperatures in Celsius per month from 2000 to 2018 measured in Berlin-Dahlem, Germany

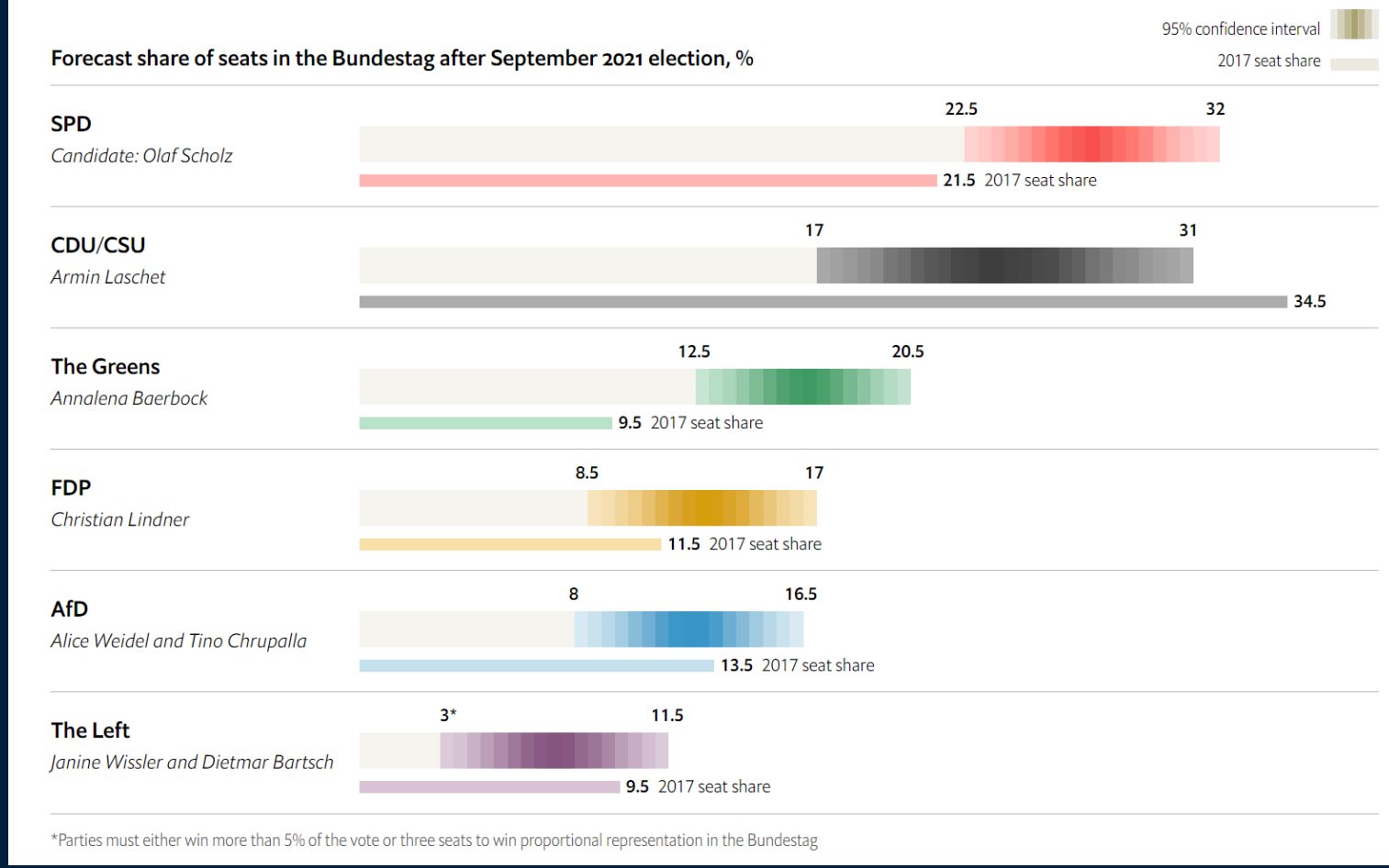


Visualization by Cédric Scherer | Data: DWD (Deutscher Wetterdienst)

GRADIENT INTERVAL STRIPS

ggdist::stat_gradientinterval()

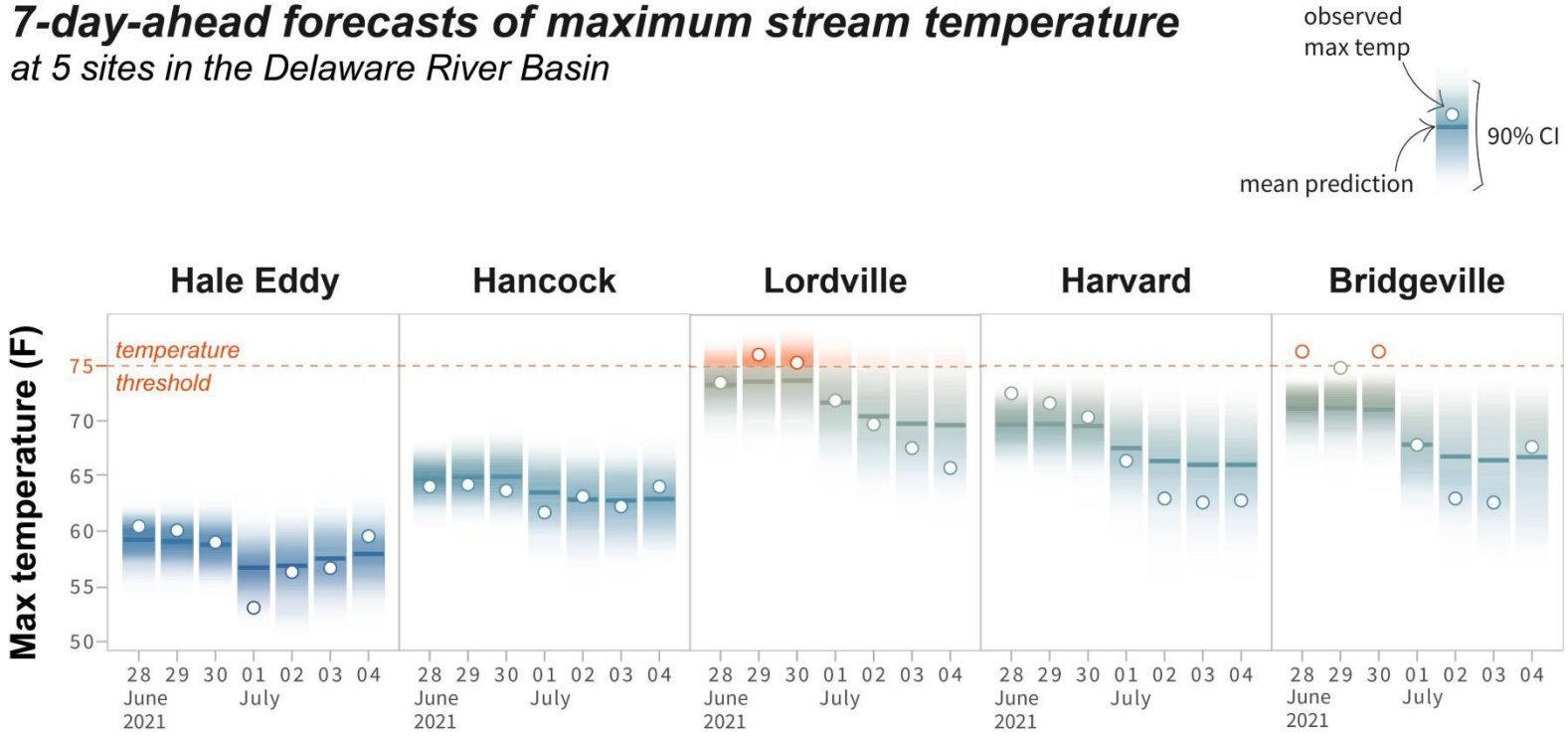




"Who will succeed Angela Merkel?" *by The Economist*

7-day-ahead forecasts of maximum stream temperature

at 5 sites in the Delaware River Basin

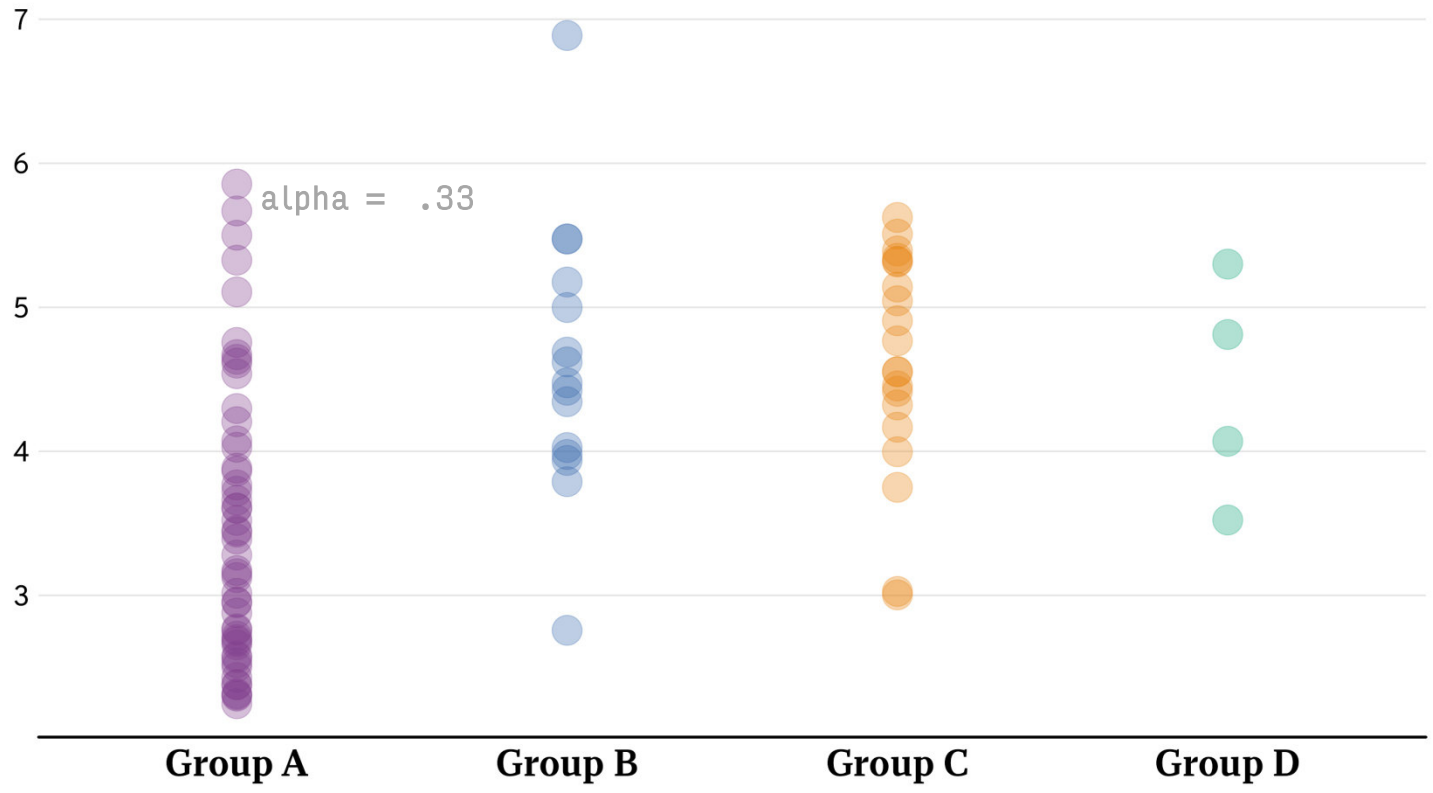


Zwart et al. (2021) Earth ArXiv (doi.org/10.31223/X55K7G)

SHOW THE *Raw* DATA

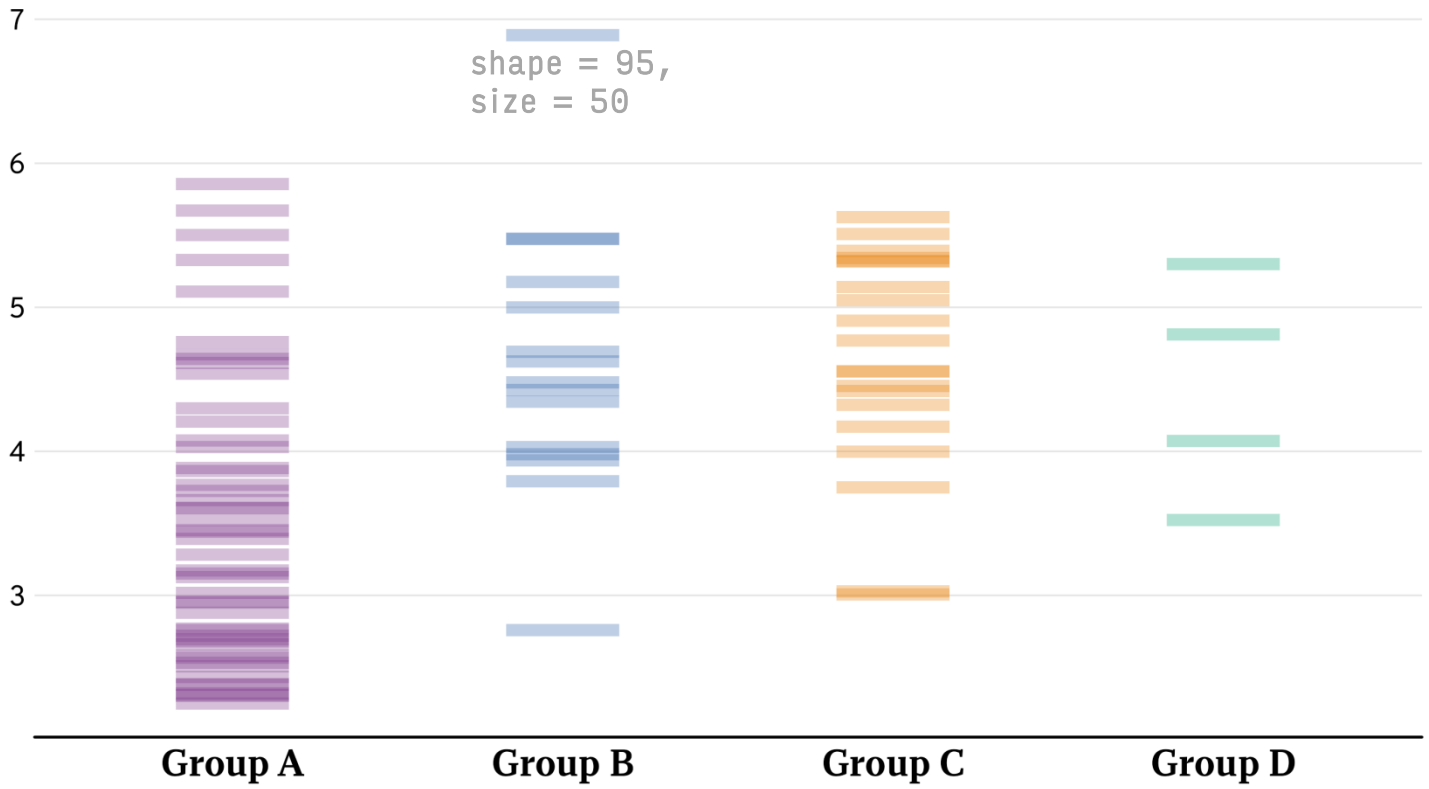
SCATTERPLOT (STRIP CHART)

geom_point()



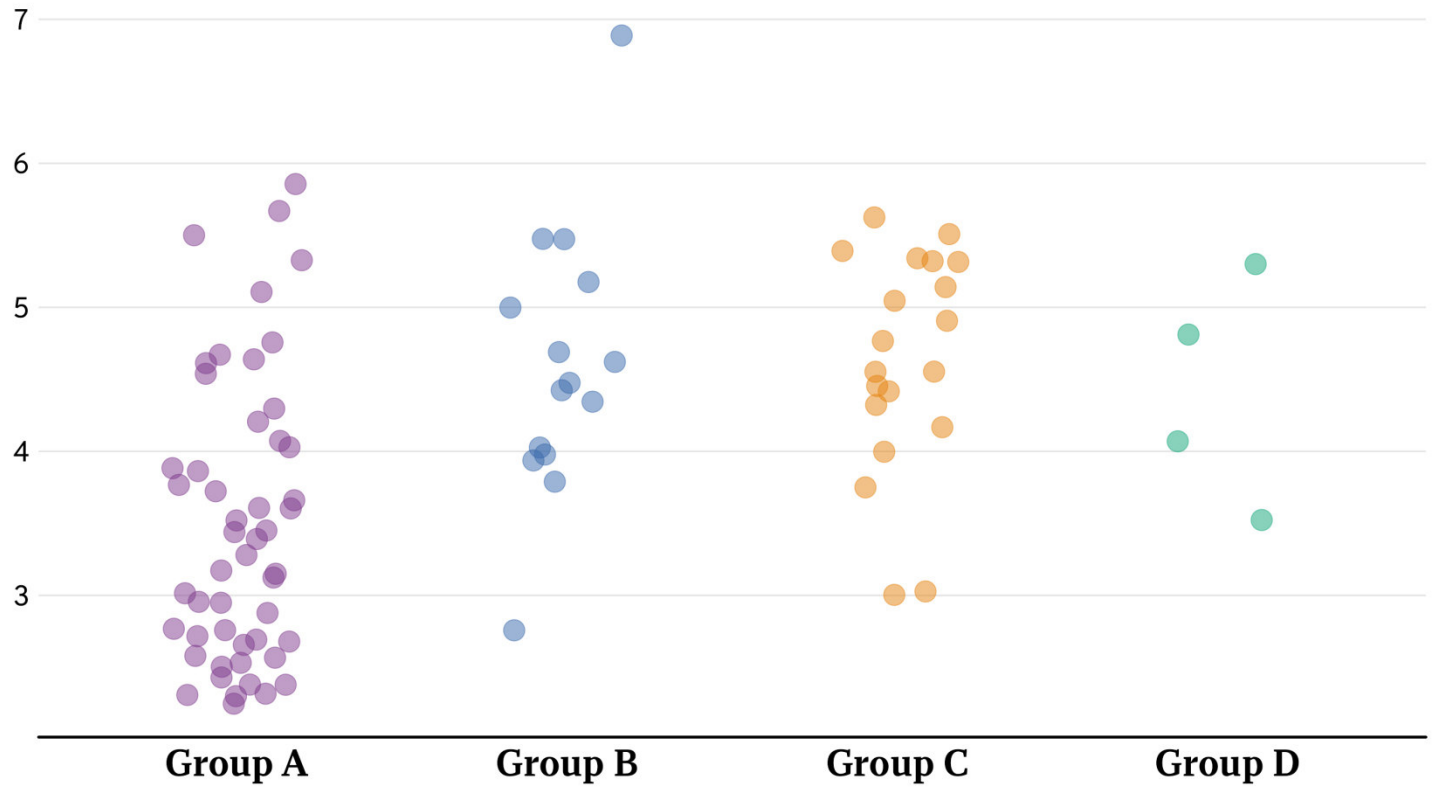
BARCODE PLOT (STRIP CHART)

geom_point()



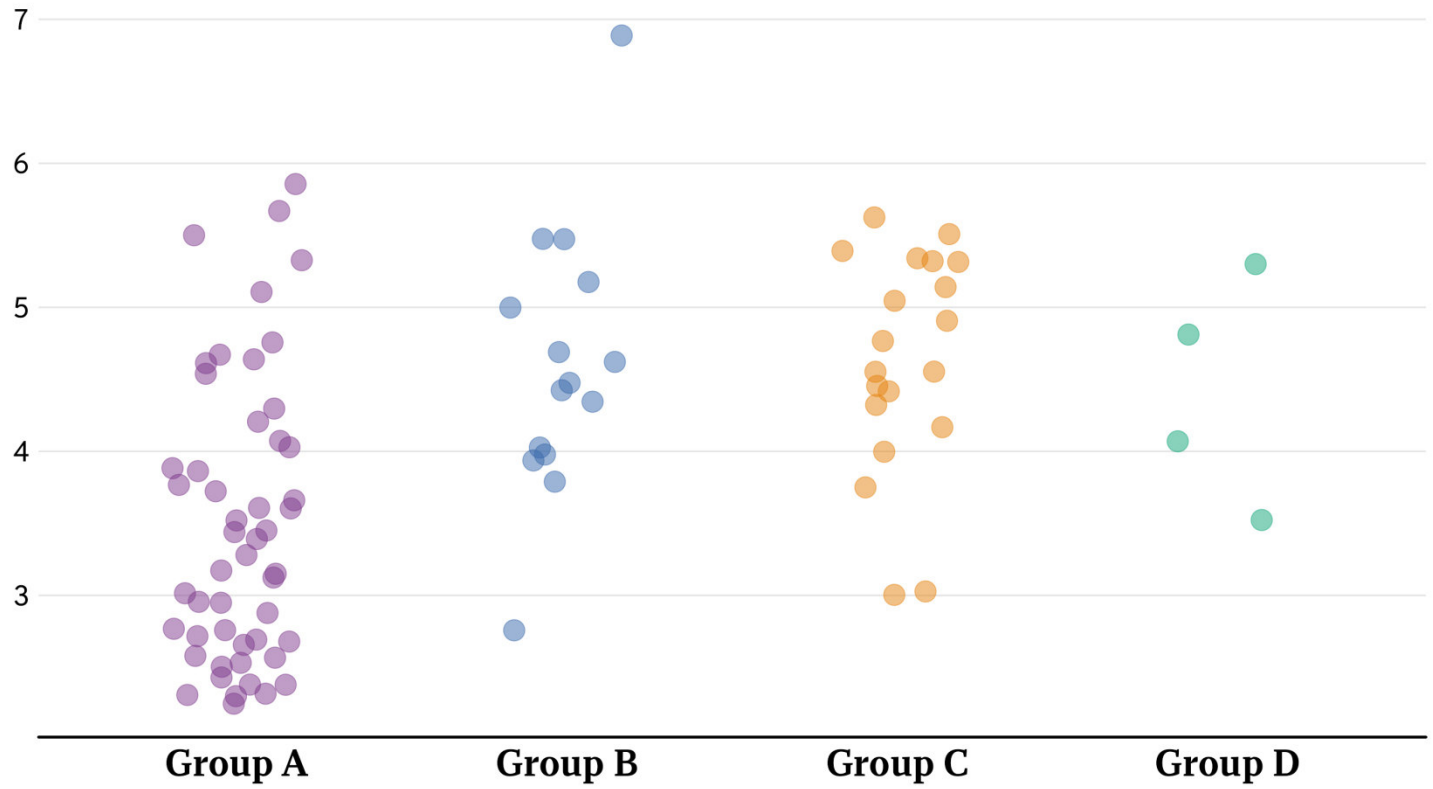
JITTER STRIP CHART

geom_jitter()



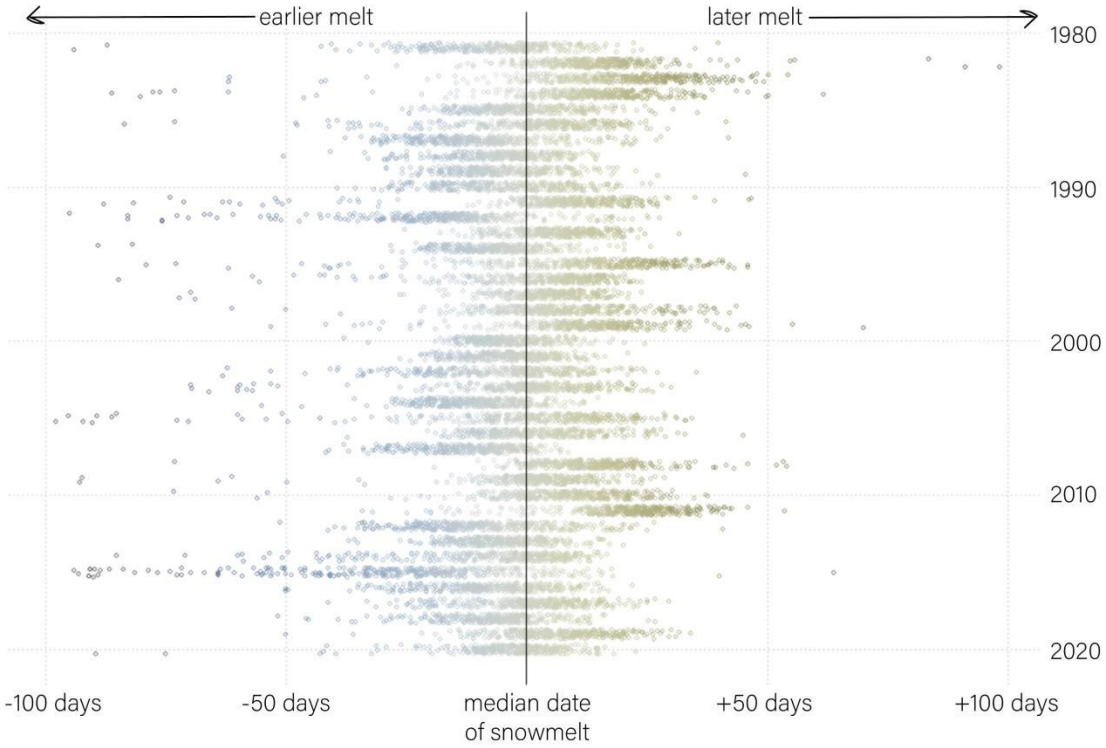
JITTER STRIP CHART

`geom_point(position = position_jitter())`



Snowmelt timing (1981-2020)

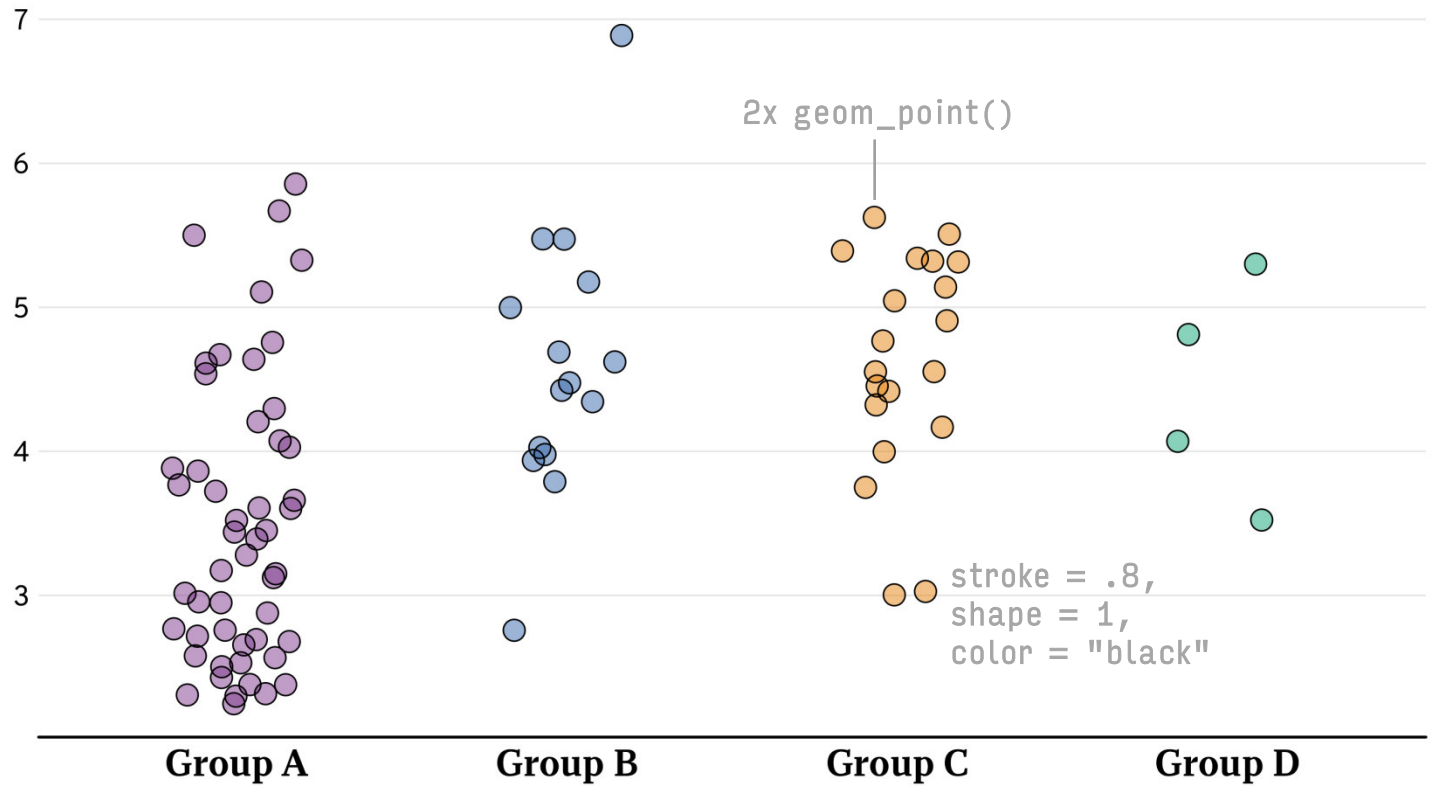
the difference in days from median snowmelt date at USDA-NRCS snow telemetry (SNOTEL) sites



Contribution to #30DayChartChallenge by @USGS_DataSci

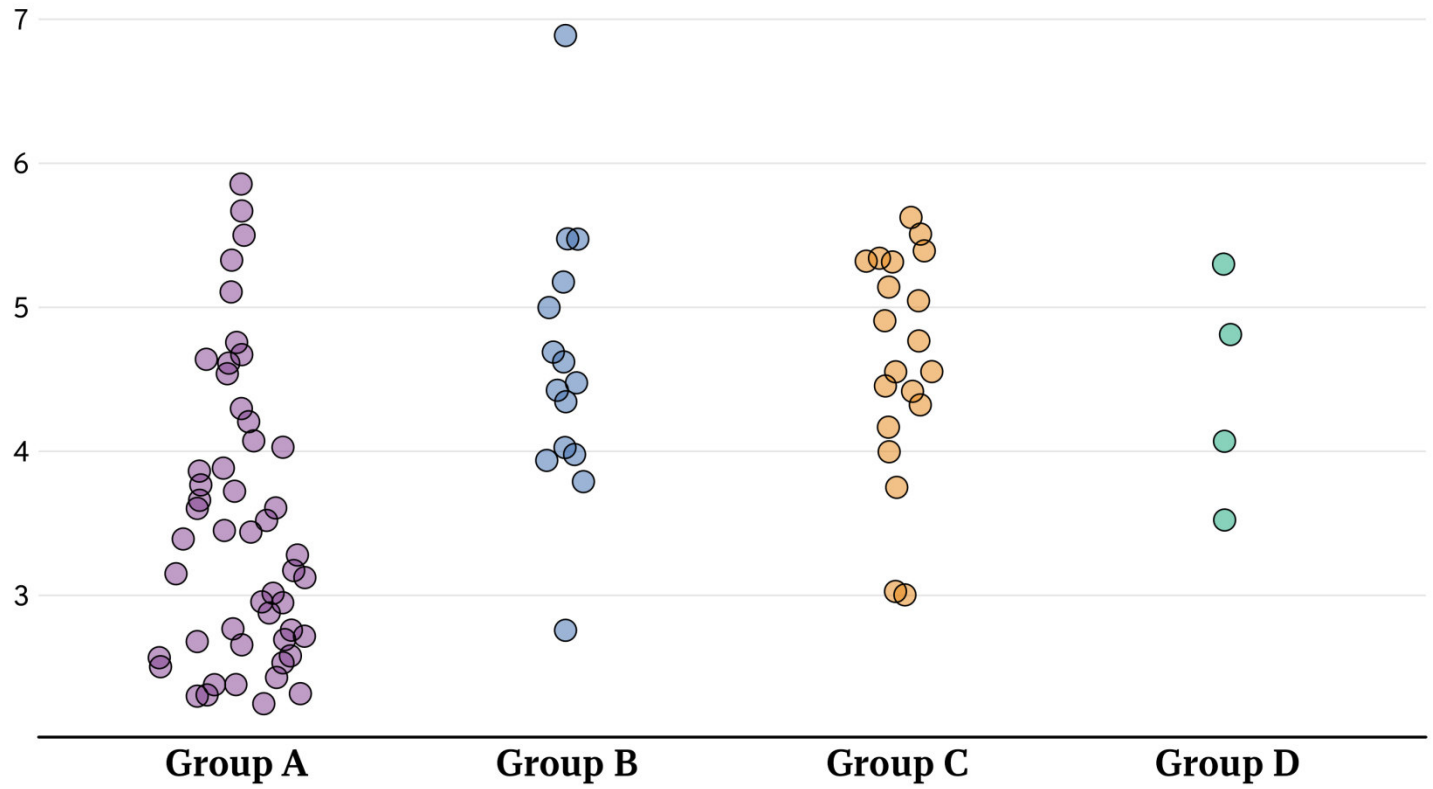
JITTER STRIP CHART

geom_point(position = position_jitter(seed = 0))



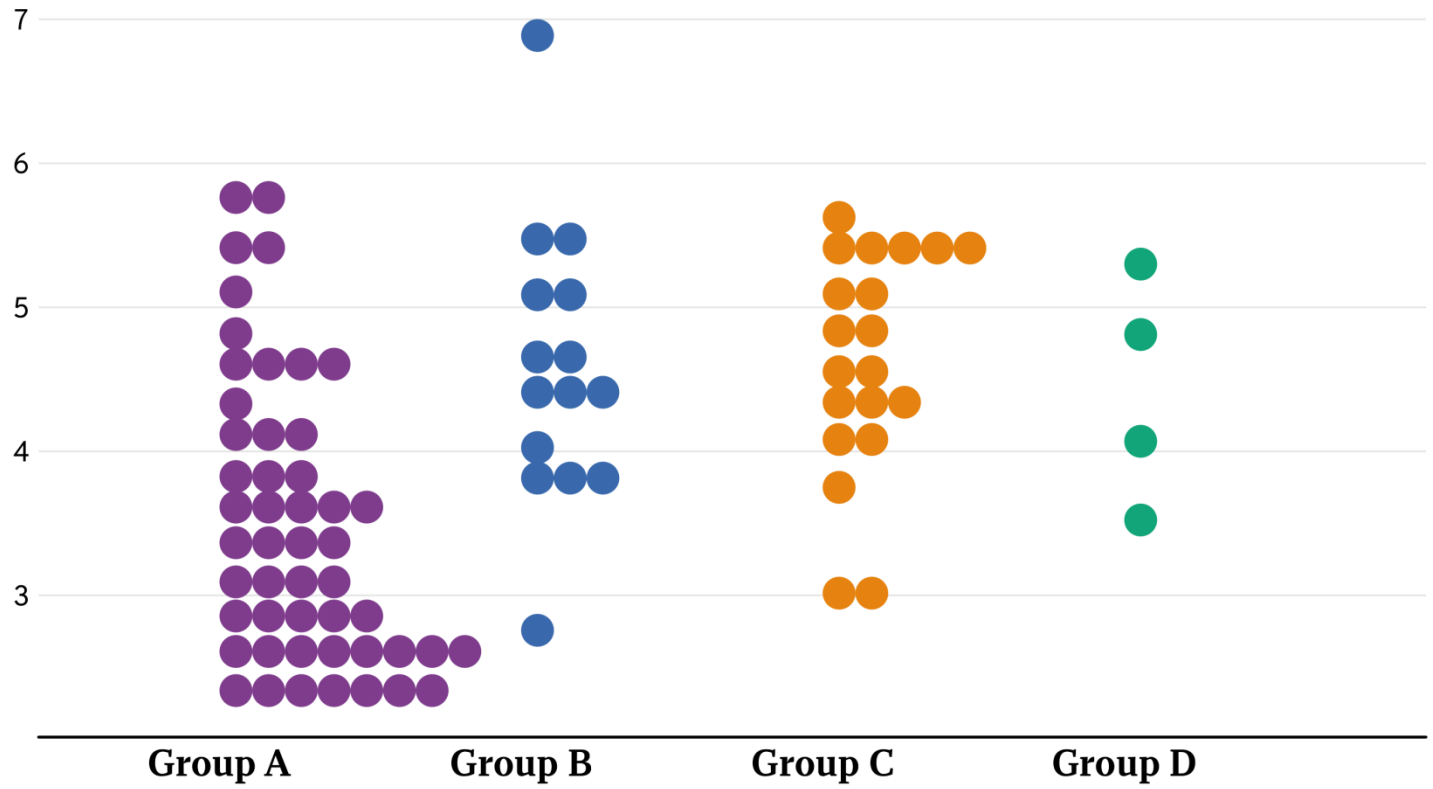
SINA PLOT

ggforce::geom_sina(seed = 0)



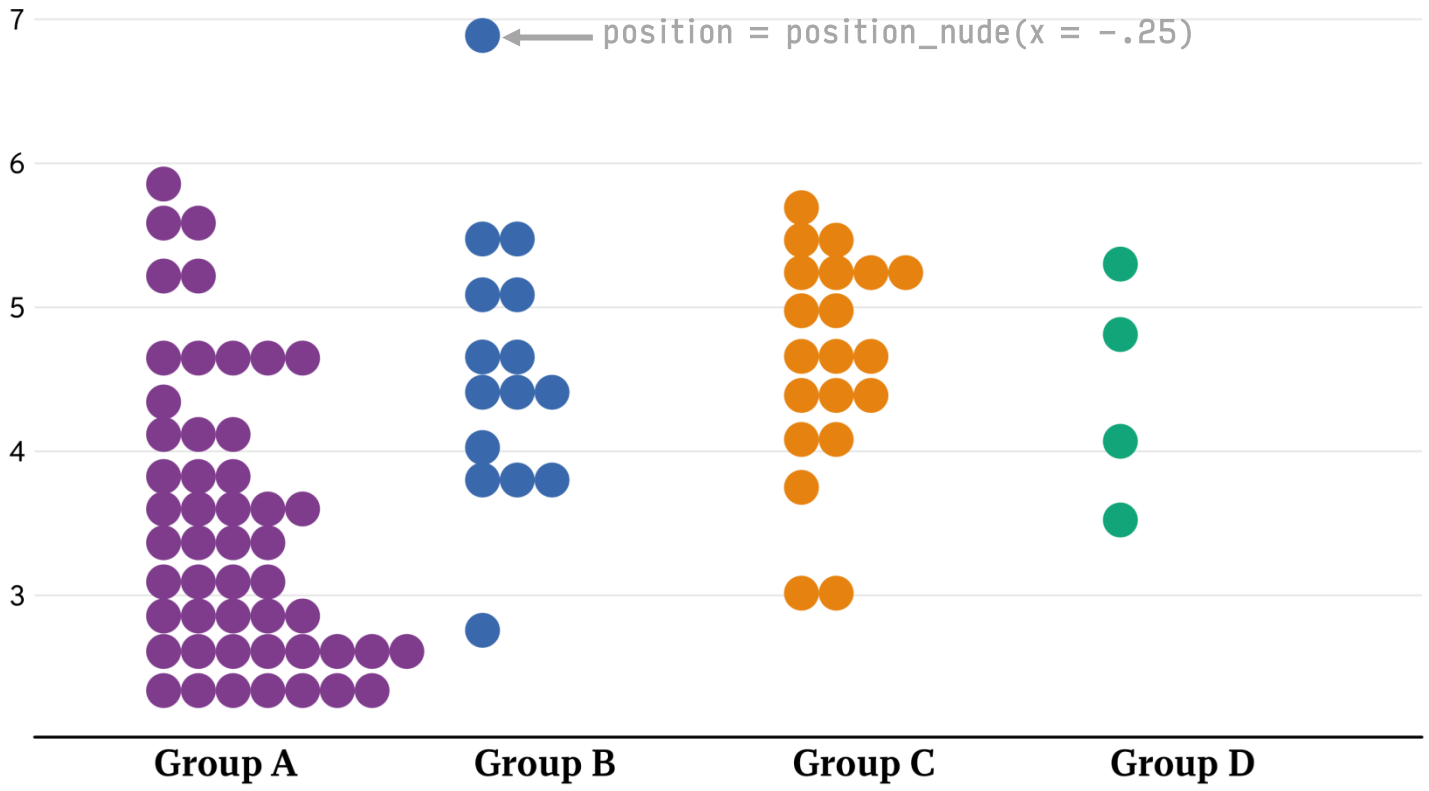
DOT PLOT

ggdist::stat_dots()



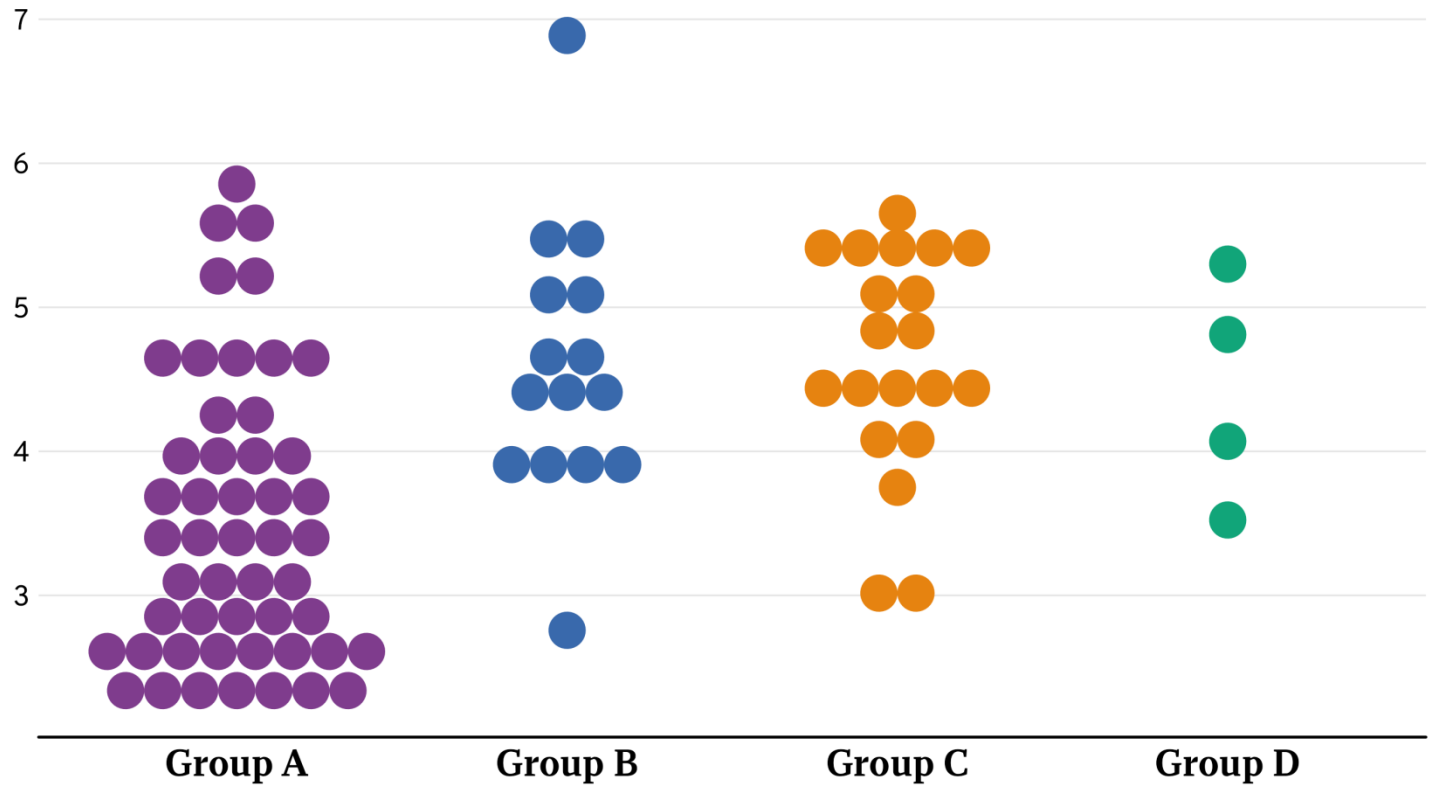
DOT PLOT

ggdist::stat_dots()



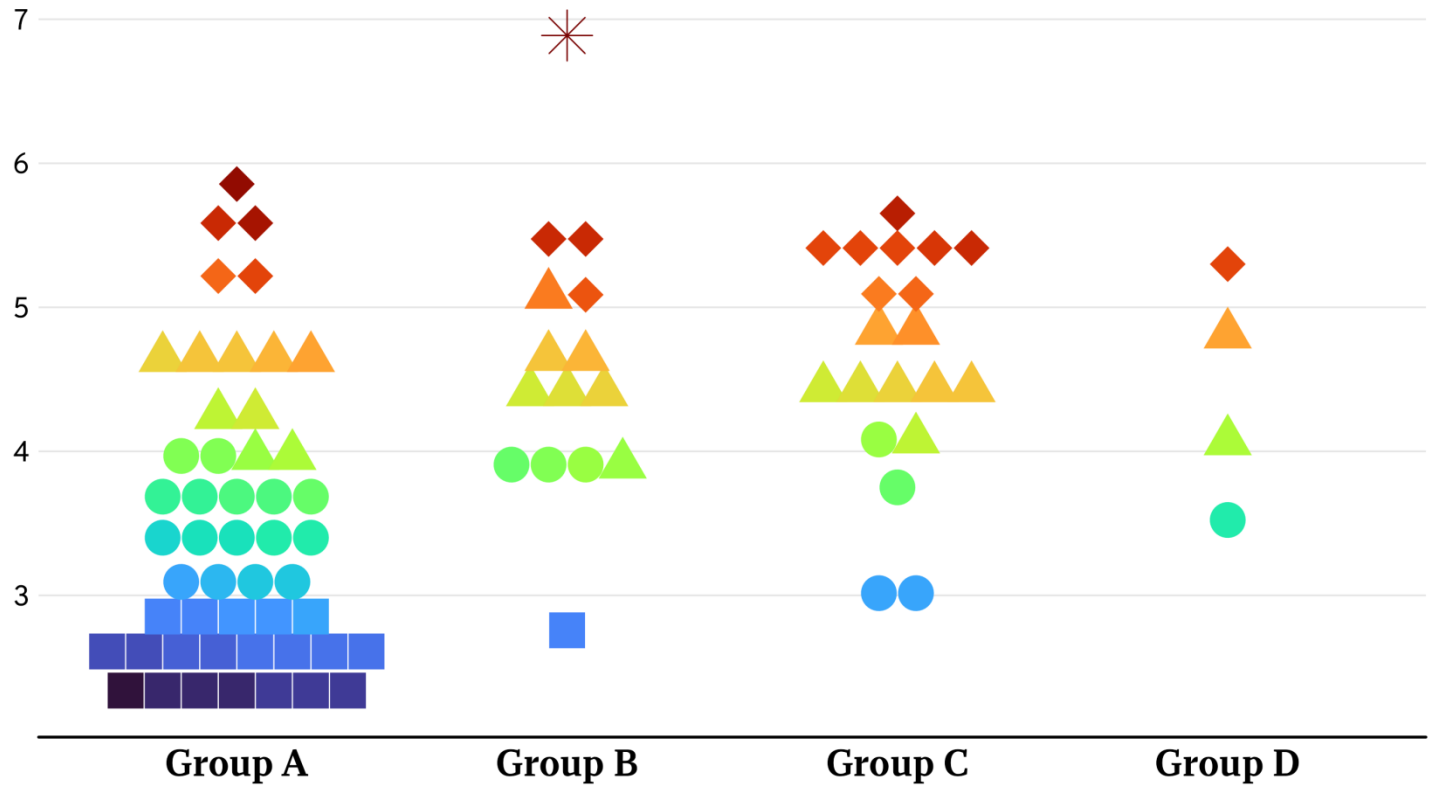
DOT PLOT

ggdist::stat_dots(side = "both")



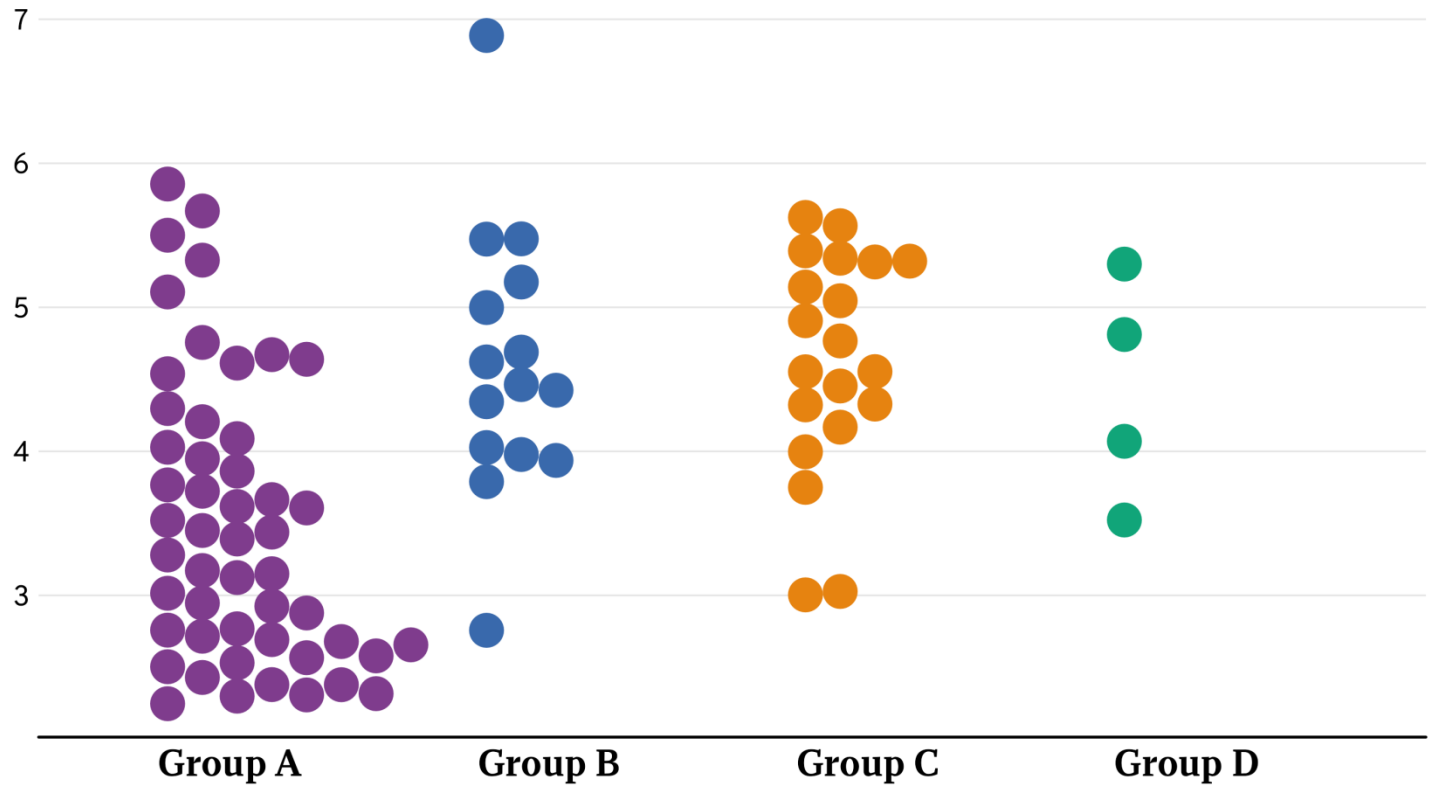
DOT PLOT

ggdist::stat_dots(side = "both")



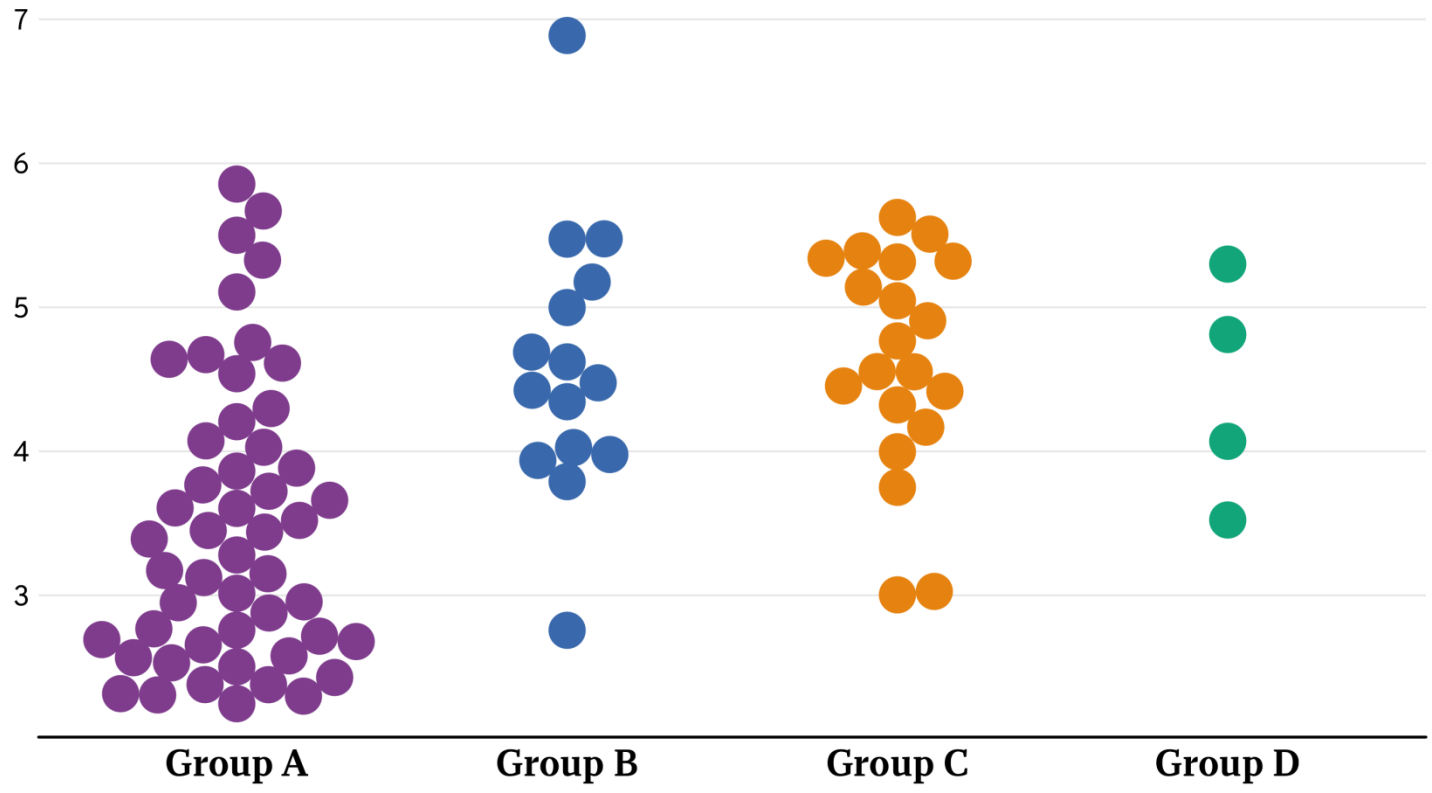
DOT PLOT

ggdist::stat_dots(layout = "weave")



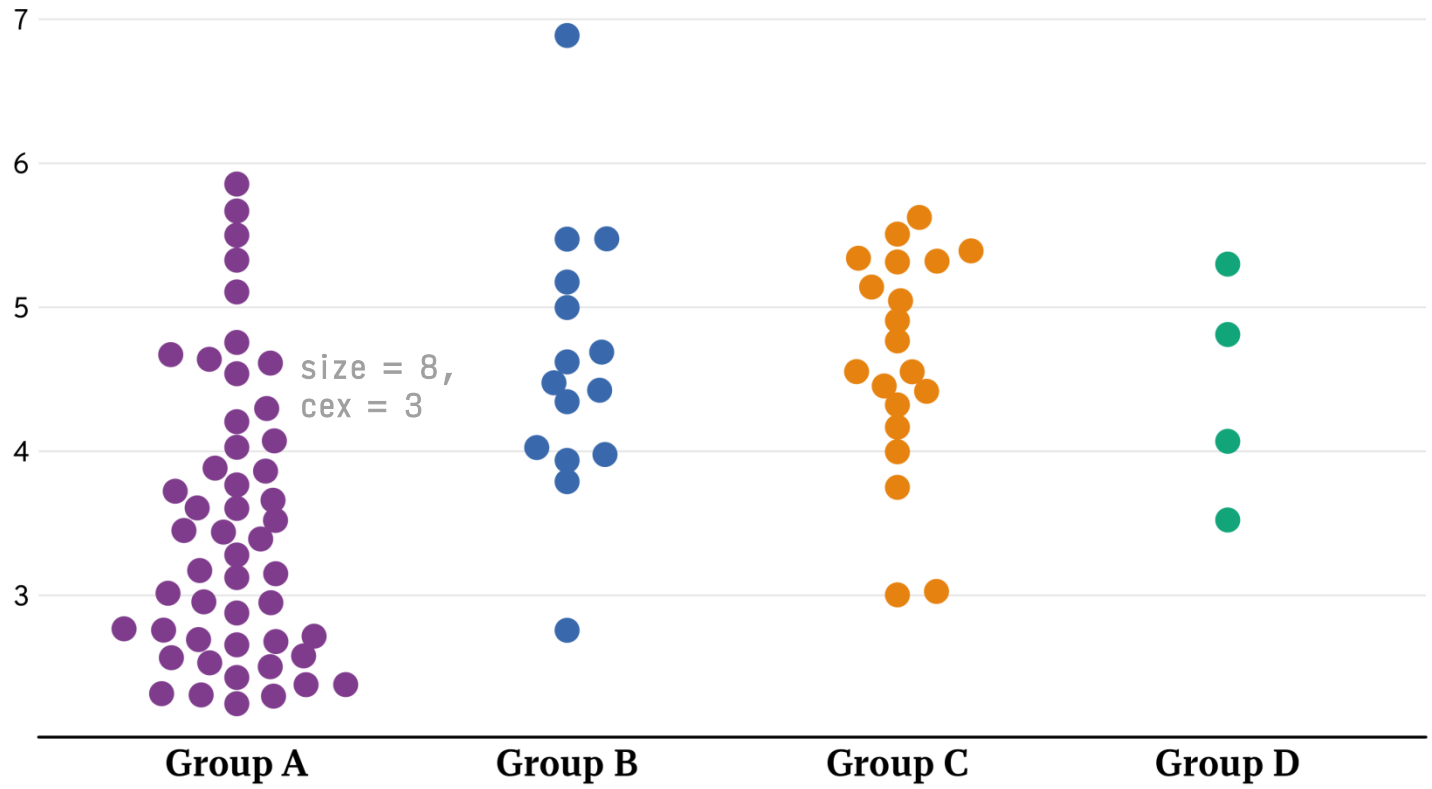
BEESWARM PLOT

ggdist::stat_dots(layout = "swarm", side = "both")



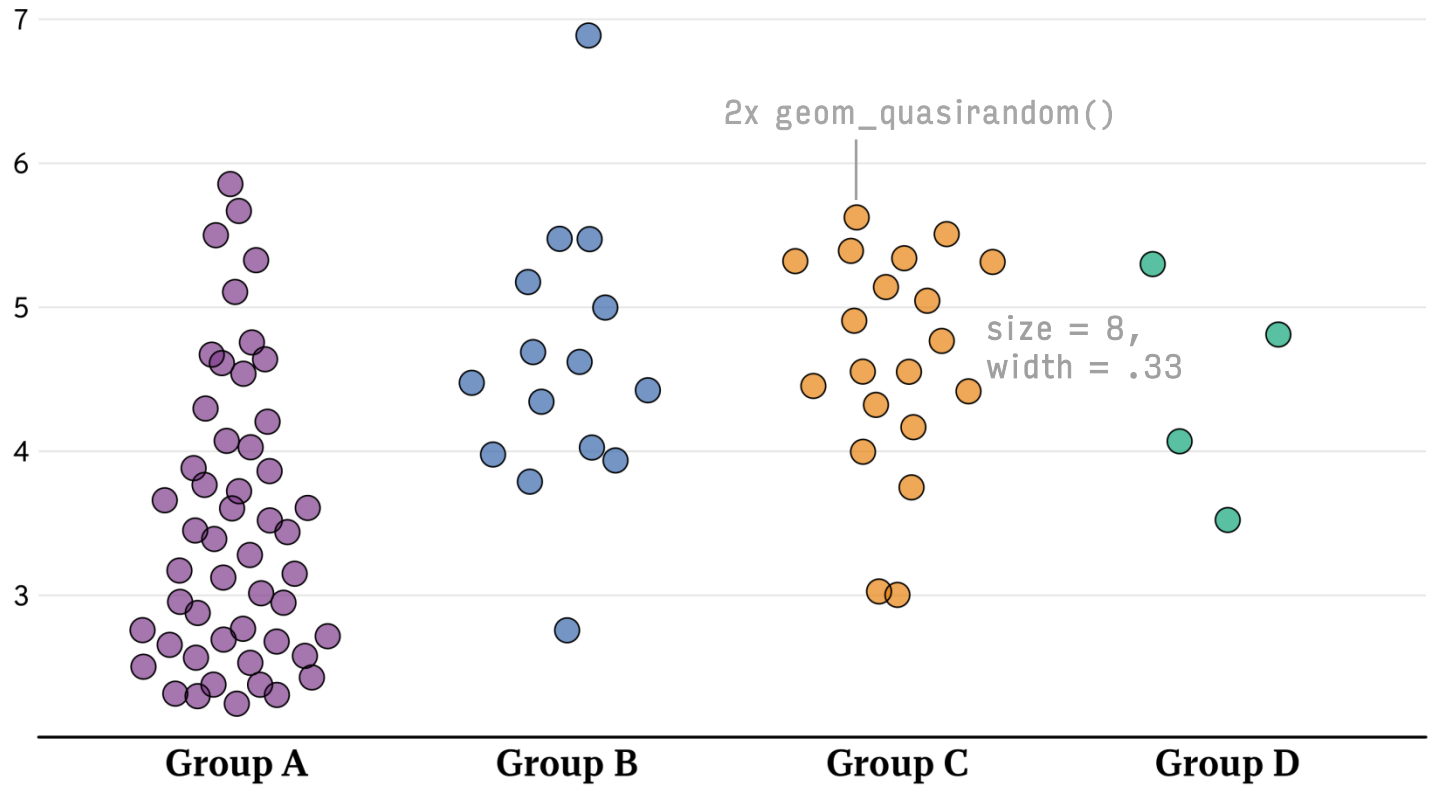
BEESWARM PLOT

ggbeeswarm::geom_beeswarm()

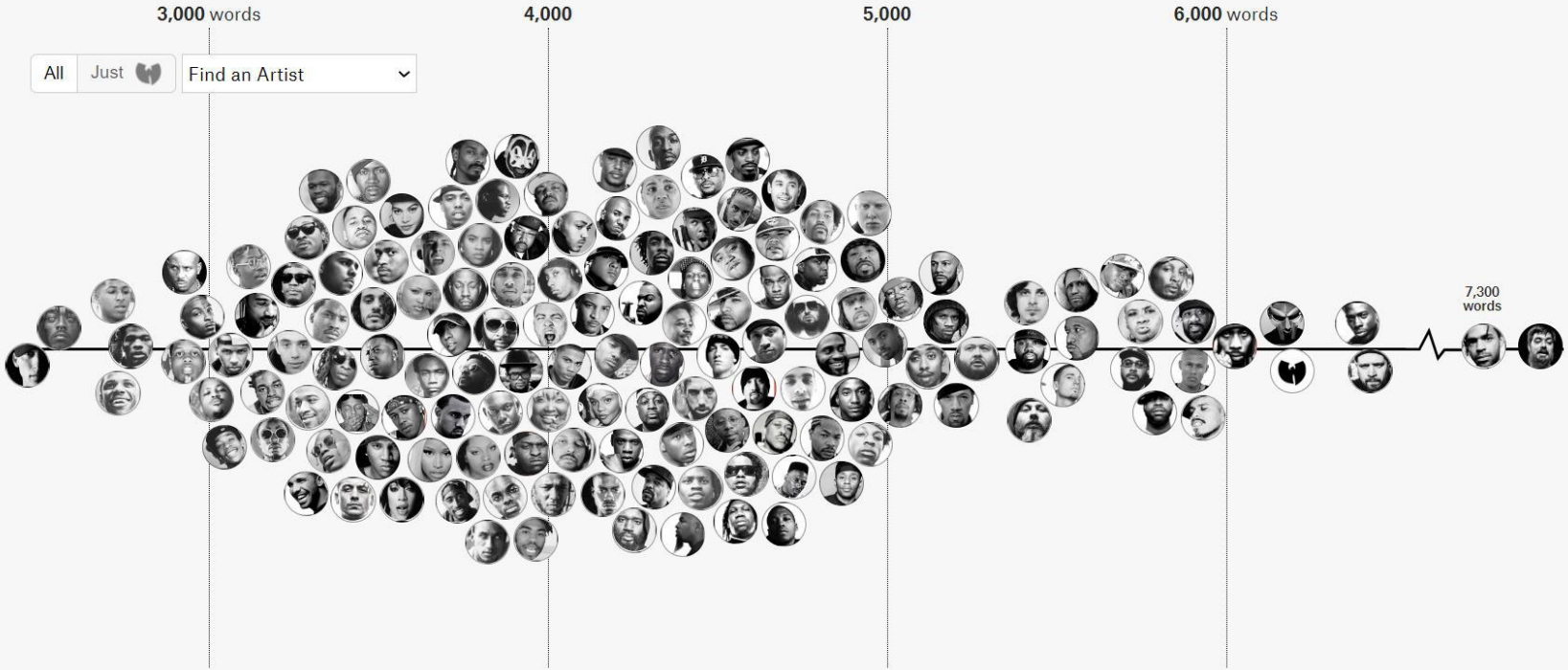


BEESWARM PLOT

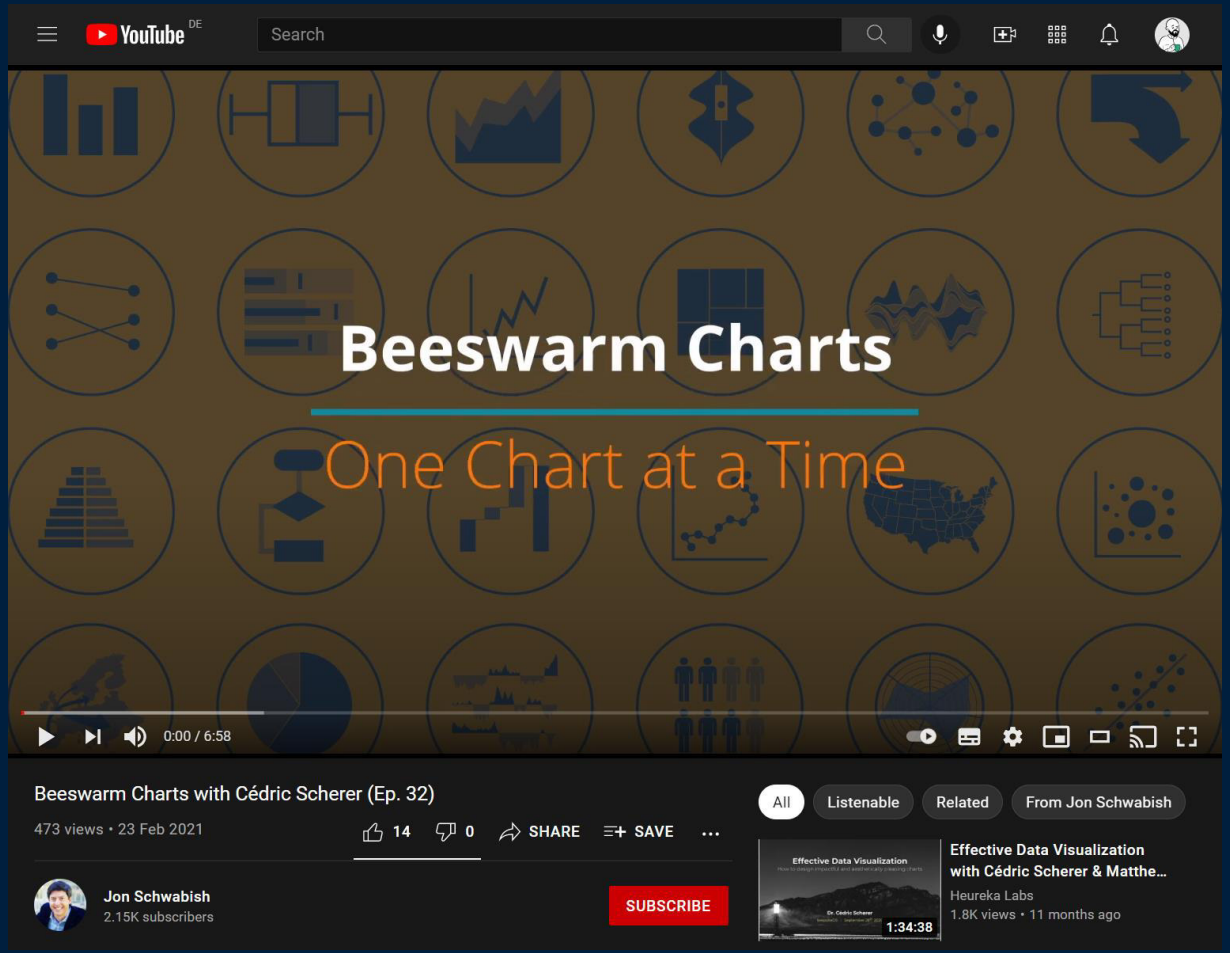
ggbeeswarm::geom_quasirandom()



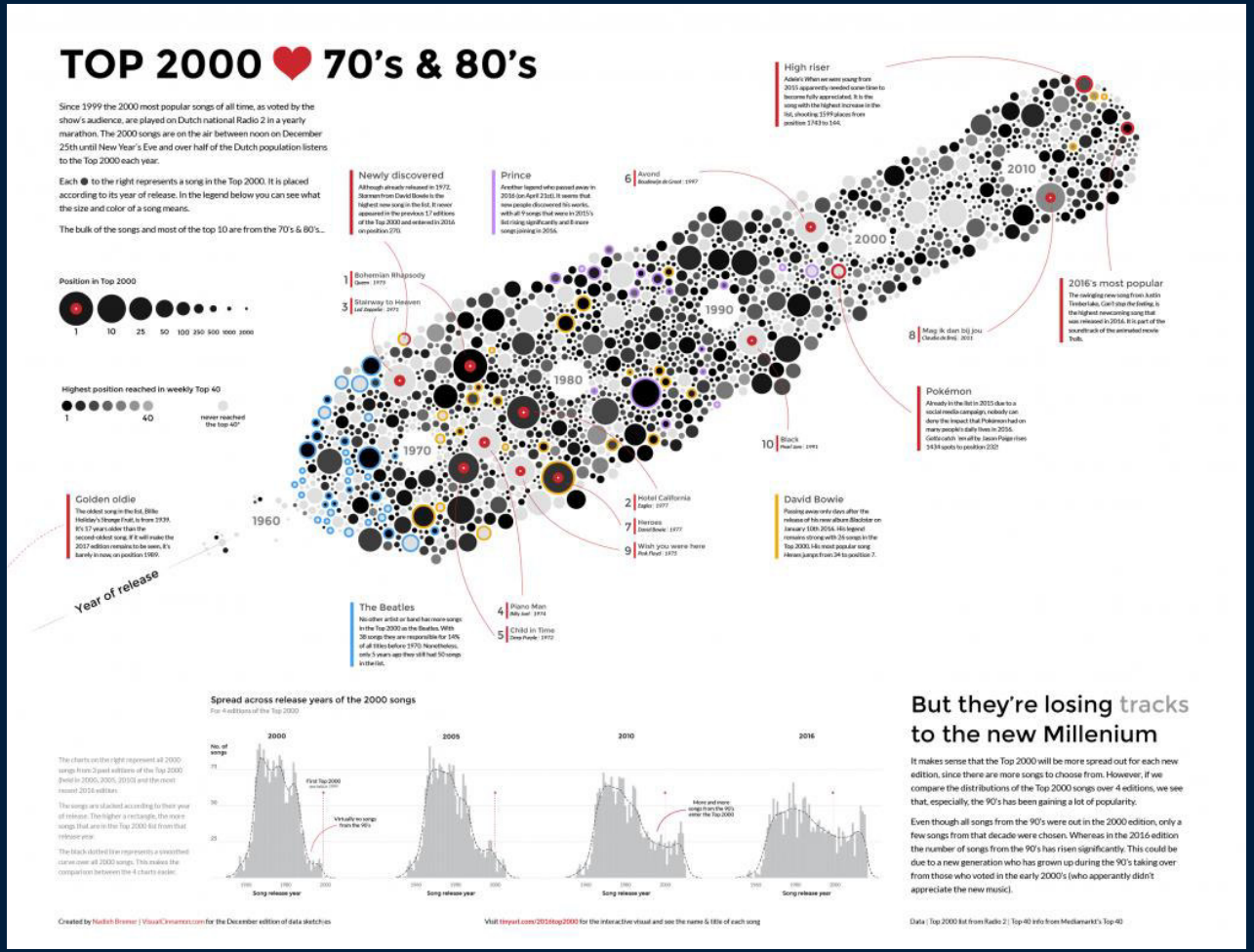
of Unique Words Used Within Artist's First 35,000 Lyrics



"The Largest Vocabulary In Hip Hop" *by Matt Daniels (The Pudding)*



youtu.be/kTtV2GRCkj0



"The Top 2000 ♥ the 70s & 80s" *by Nadieh Bremer*

How the growing racial diversity in the US is reflected in schools

Comparison of the Simpson diversity index, a quantitative measure that reflects the racial diversity, during the school years **1994–1995** and **2016–2017** for schools with the racial **diversity being lower than (• or •) or higher than/equal to (• or •) the state's median.**
Dotted lines indicate the US median diversity and numbers the number of schools with only one ethnicity (dots removed for better readability).

Visualization by Cédric Scherer | Data: NCES


Nebraska
2
348
Iowa
57
Illinois
1
114
Indiana
9
Ohio
2
8
Virginia
1
District of Columbia
Delaware
0 .2 .4 .6 .8
Kansas
14
Missouri
17
125
Tennessee
6
Kentucky
5
West Virginia
North Carolina
Maryland
0 .2 .4 .6 .8
Oklahoma
7
Arkansas
37
Mississippi
2
Alabama
Georgia
2
South Carolina
0 .2 .4 .6 .8 0 .2 .4 .6 .8 0 .2 .4 .6 .8
Texas
1
6
Louisiana
Florida
0 .2 .4 .6 .8 0 .2 .4 .6 .8 0 .2 .4 .6 .8

Leland Wilkinson
May 28, 2016 at 3:43 pm

People interested in these plots need to read my paper

http://moderngraphics11.pbworks.com/f/wilkinson_1999.DotPlots.pdf

before programming dot plots. Some of the implementations above are useful. One, however, is misleading and should be avoided. That's the "beeswarm" plot (why did we need a new name for this 100 year old plot?). In the vertical version of the "beeswarm" plot, the Y values are placed at their proper locations but the X values are arbitrarily ordered by the Y values. This creates a visual artifact of U-shaped dot stacks that misrepresent the structure of the data. There are also other examples in the "beeswarm" R program that allow the dots to be asymmetric around a vertical center line. This, too, induces a visual artifact. Dot plots need to be a faithful representation of a density (this is a well-defined statistical concept) and need to converge to a population density as sample size increases.

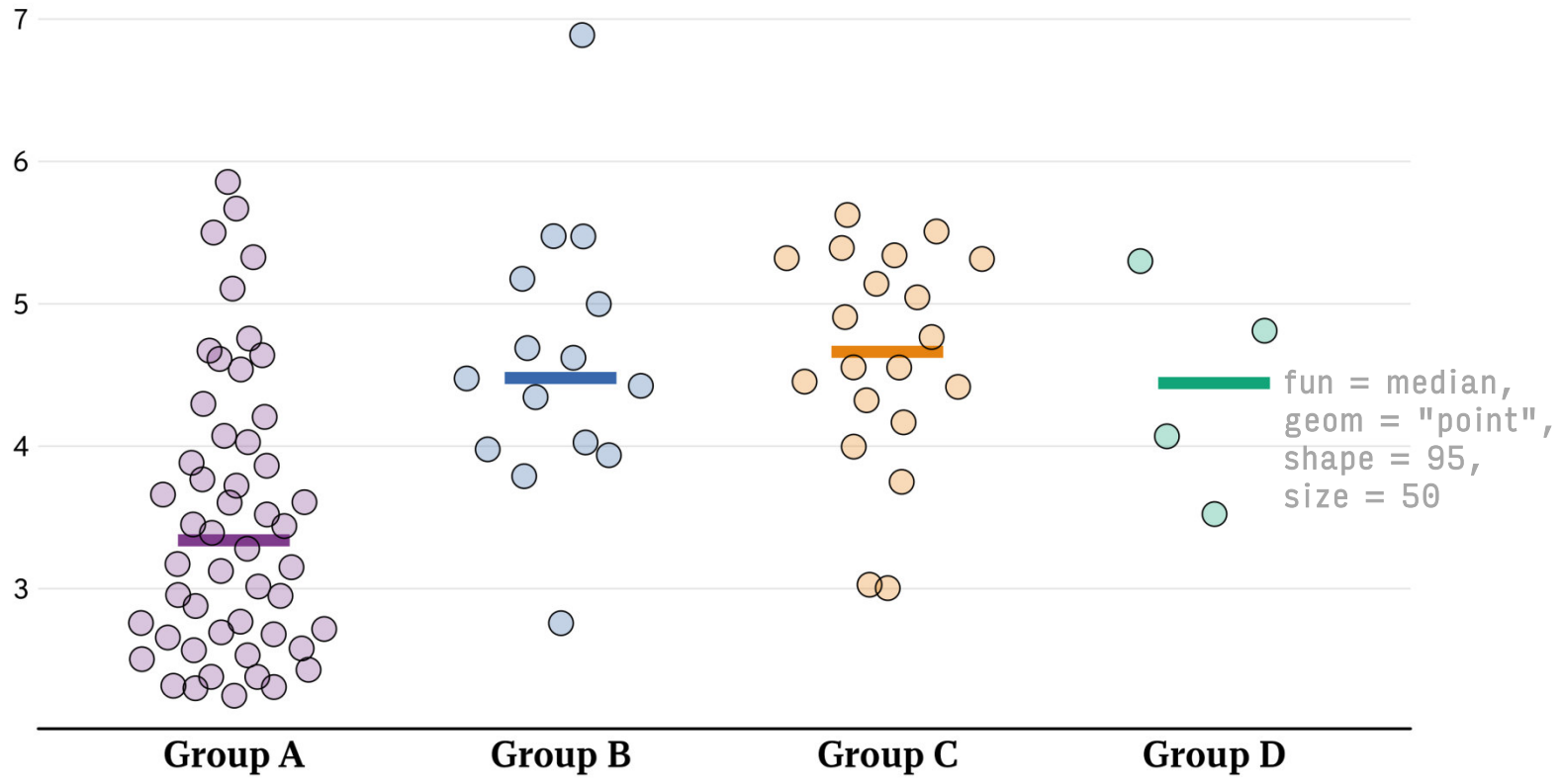
The main point of my paper was not to devise an algorithm for producing dot plots, but to show that an admissible dot plot algorithm (there are several) needs to be evaluated on its Integrated Mean Square Error (IMSE). I also showed that when this is properly defined, the dot plot loss function resembles the ones used for histograms and kernel density estimates.

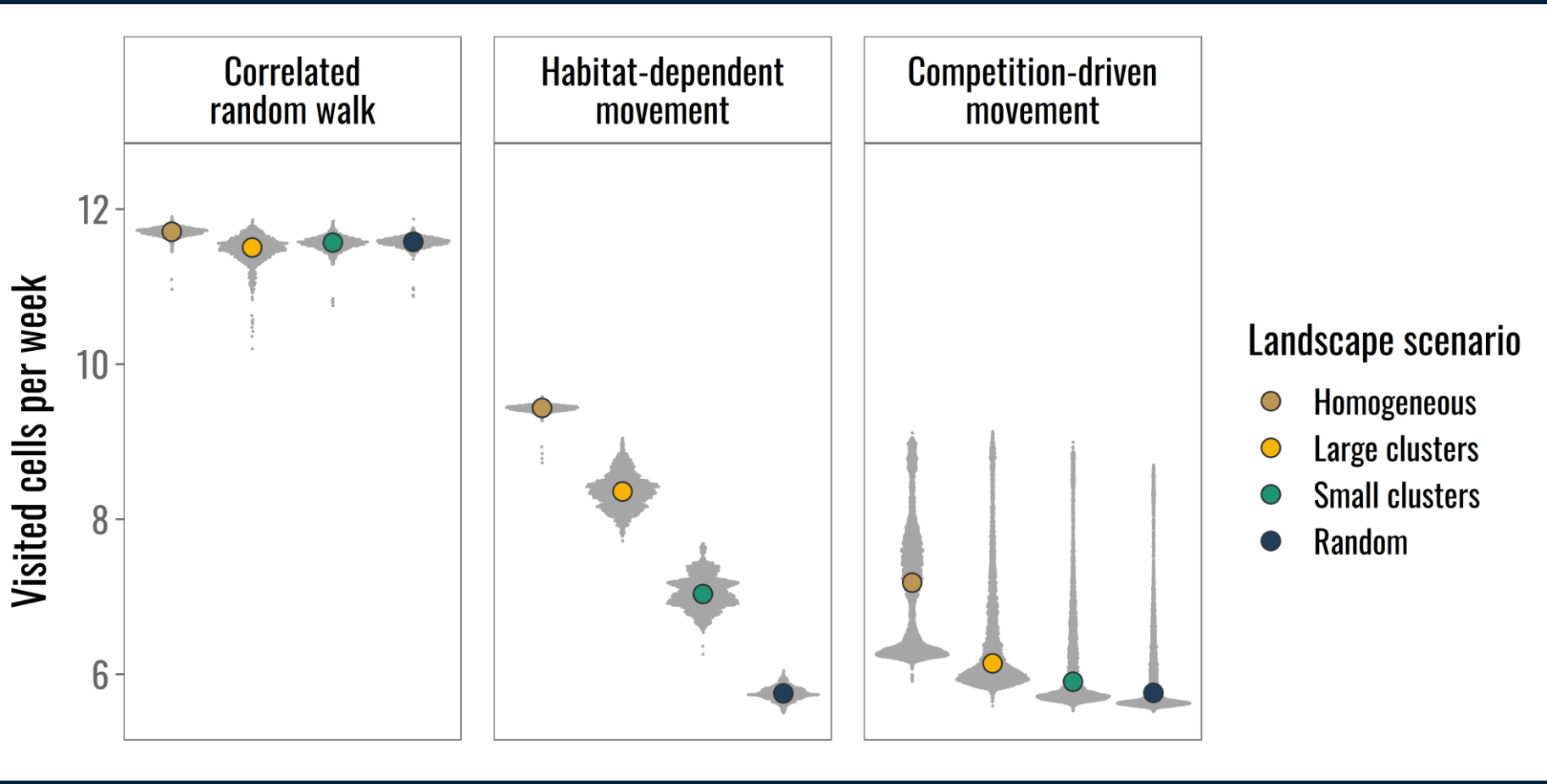
As I tried to explain in The Grammar of Graphics, valid visualizations need to pay attention to the underlying mathematical and statistical models on which they are based. It is not enough to draw unstructured pictures of "data," pretty as they may be.

THE BEST OF BOTH WORLDS:
Hybrid Charts

BEESWARM PLOT WITH MEDIAN

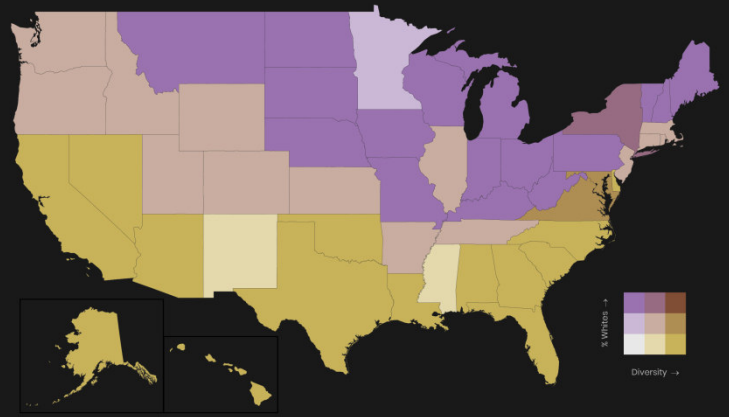
ggbeeswarm::geom_beeswarm() + stat_summary()



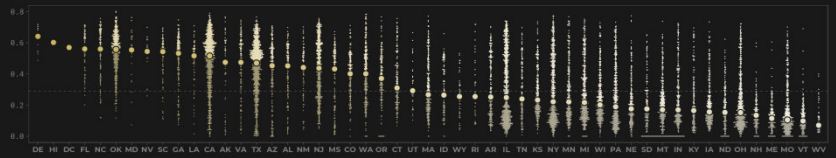

Correlated random walk
Habitat-dependent movement
Competition-driven movement
Visited cells per week
12
10
8
6
Landscape scenario
Homogeneous
Large clusters
Small clusters
Random

How diverse are schools in the US?

Bivariate map showing the combination of racial diversity measured as **Simpson index**, a quantitative diversity measure, and the **proportion of students with white ethnicity** during the **school year 2016–2017**.



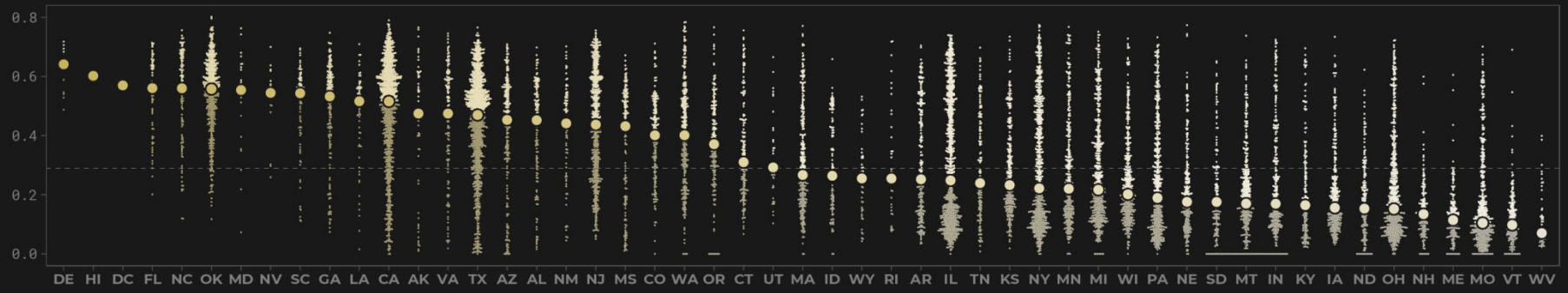
Simpson diversity index for all schools grouped per state, ranked from states with **high diversity** to those with **low diversity**.
Larger dots represent each state's median diversity with darker colored points laying below and lighter colored points laying above this value.



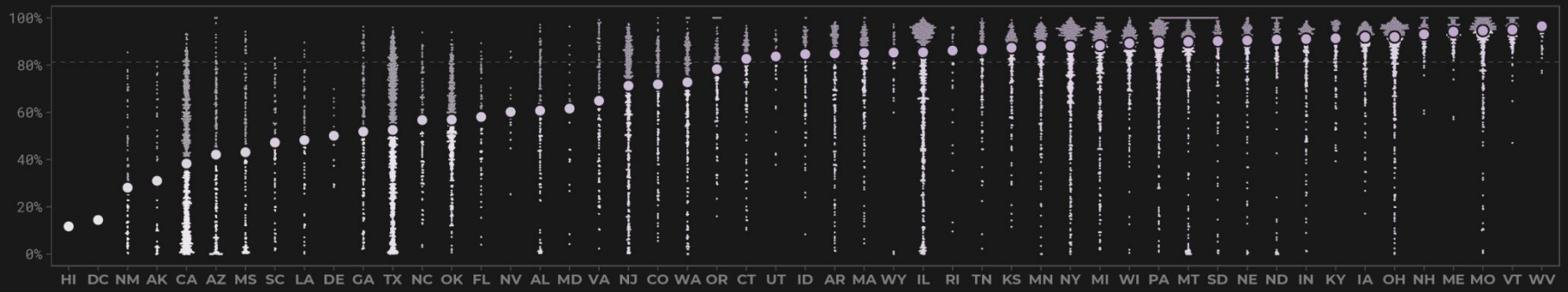
Proportion of white students for all schools grouped per state, ranked from states with **few white students** to those with **many white students**.
Larger dots represent each state's median proportion with lighter colored points laying below and darker colored points laying above this value.

Visualization by Cédric Scherer | Data: NCES

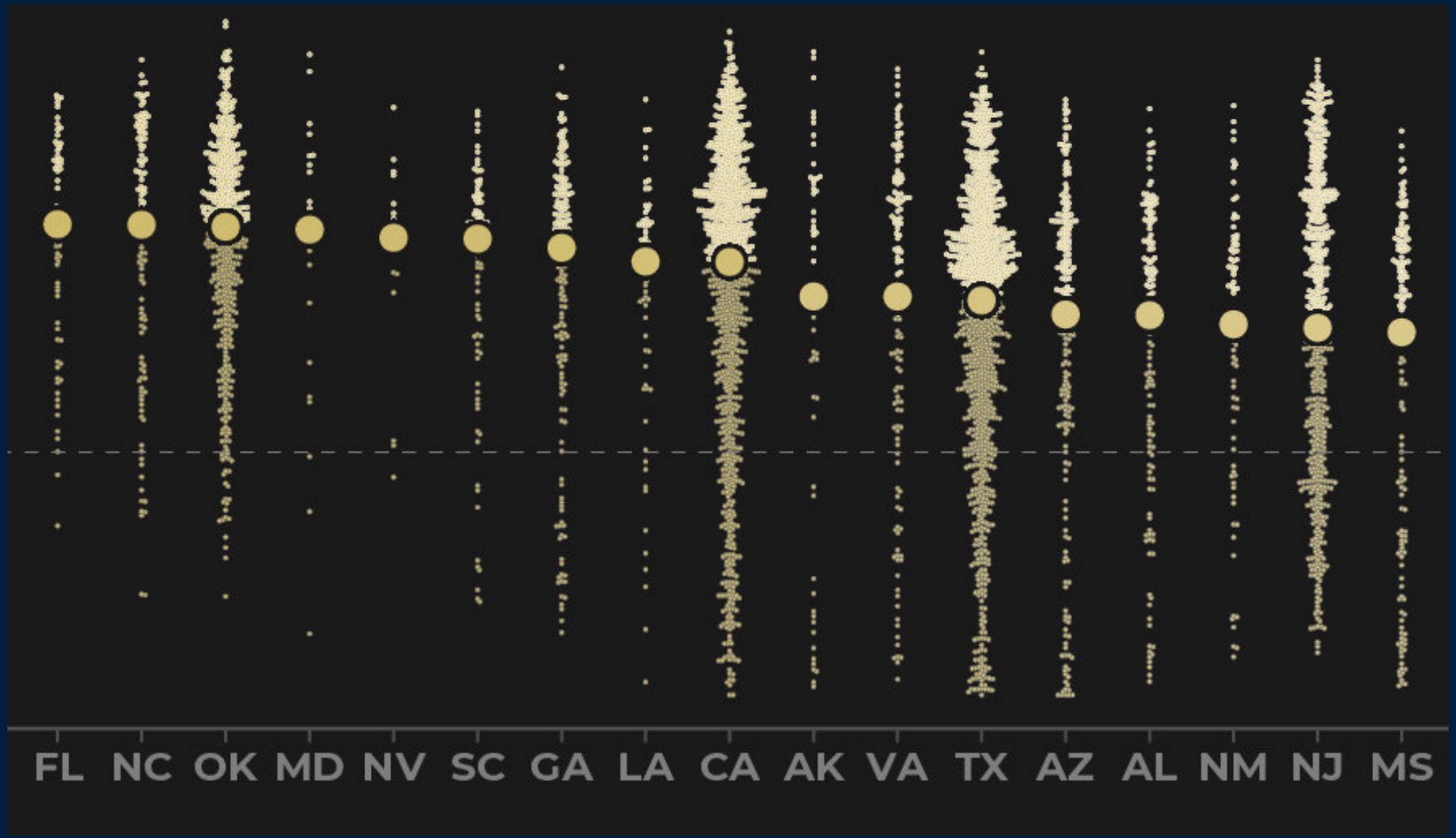
Simpson diversity index for all schools grouped per state, ranked from states with **high diversity** to those with **low diversity**.
Larger dots represent each state's median diversity with darker colored points laying below and lighter colored points laying above this value.



Proportion of white students for all schools grouped per state, ranked from states with **few white students** to those with **many white students**.
Larger dots represent each state's median proportion with lighter colored points laying below and darker colored points laying above this value.

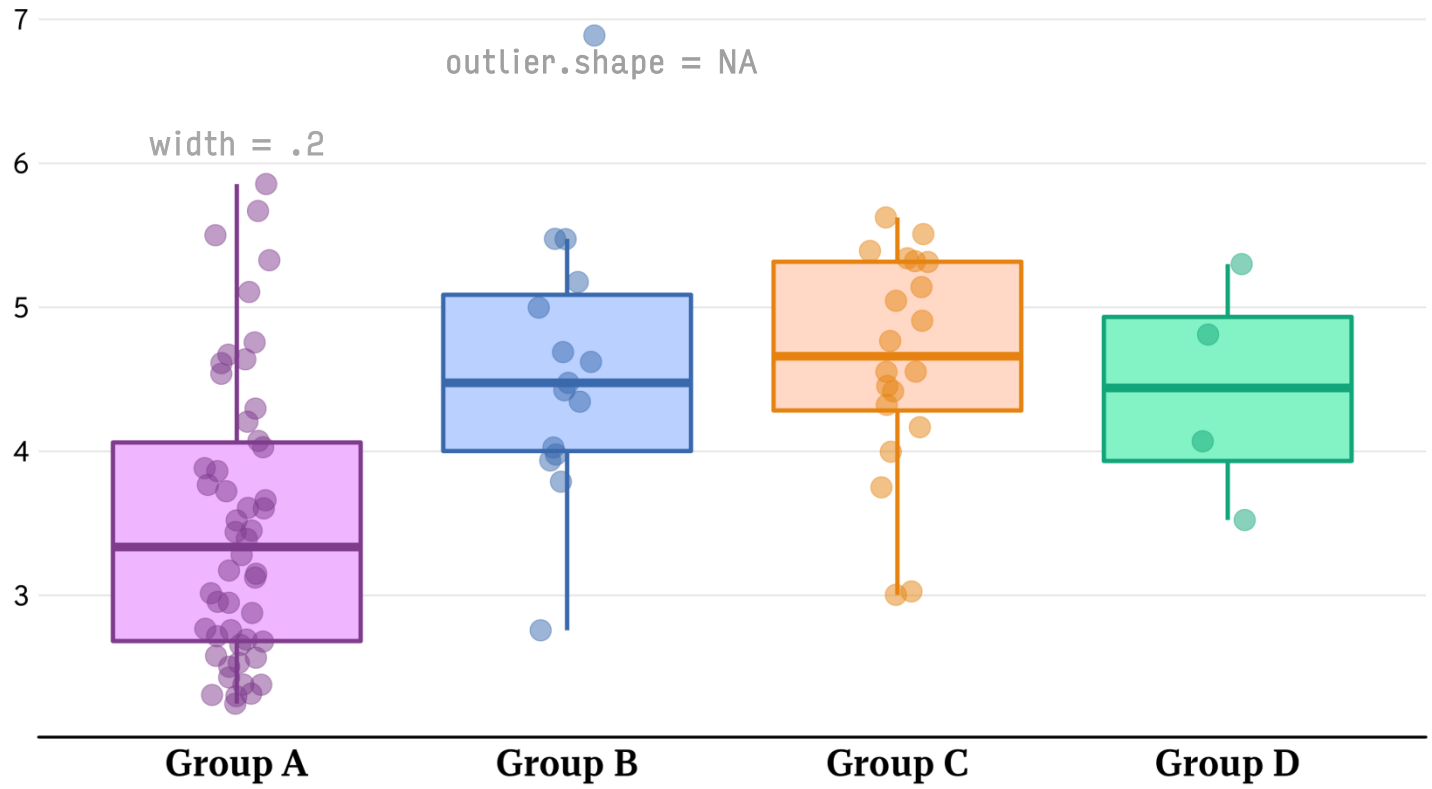


Visualization by Cédric Scherer | Data: NCES


FL
NC
OK
MD
NV
SC
GA
LA
CA
AK
VA
TX
AZ
AL
NM
NJ
MS

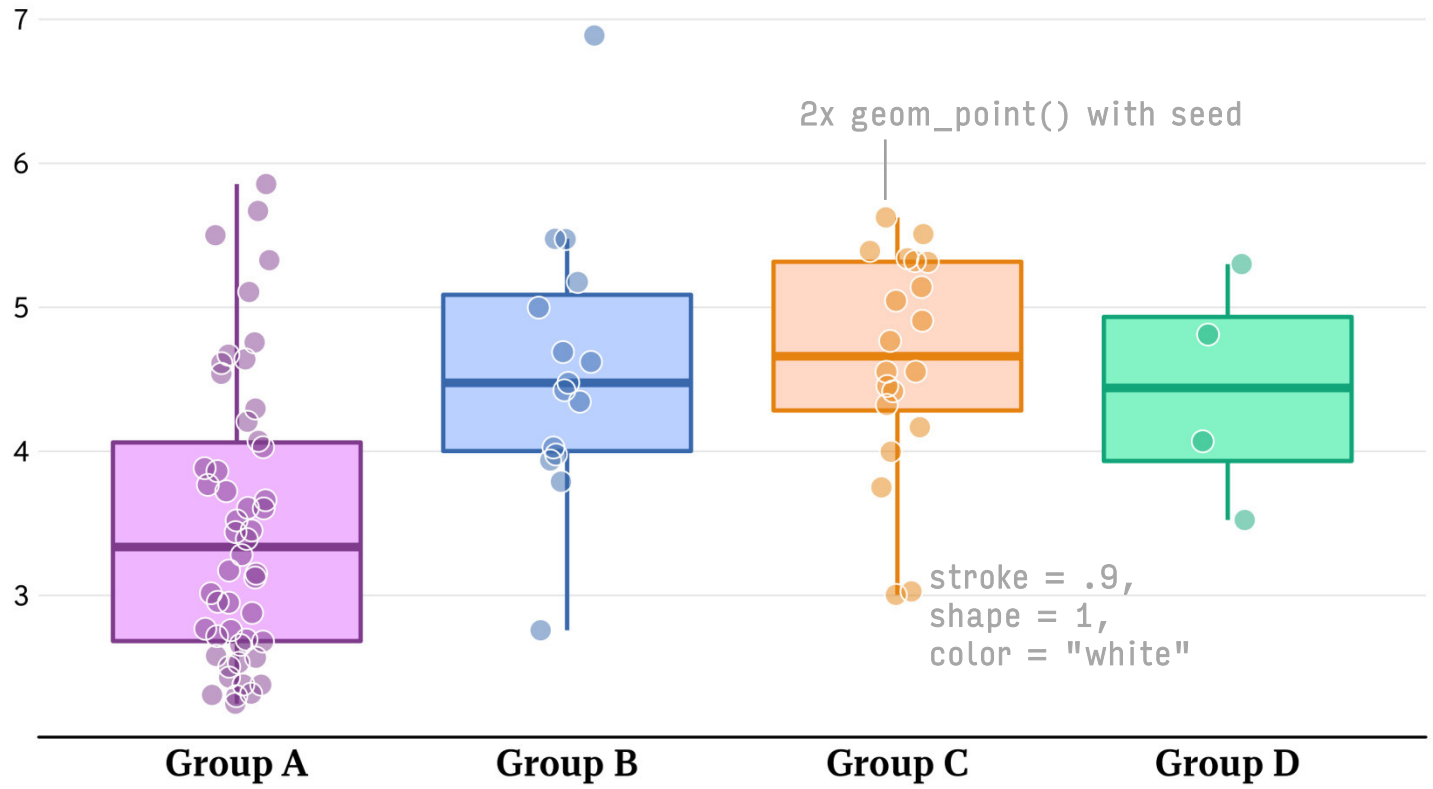
BOX PLOT X JITTER STRIPS

geom_boxplot() + geom_jitter()



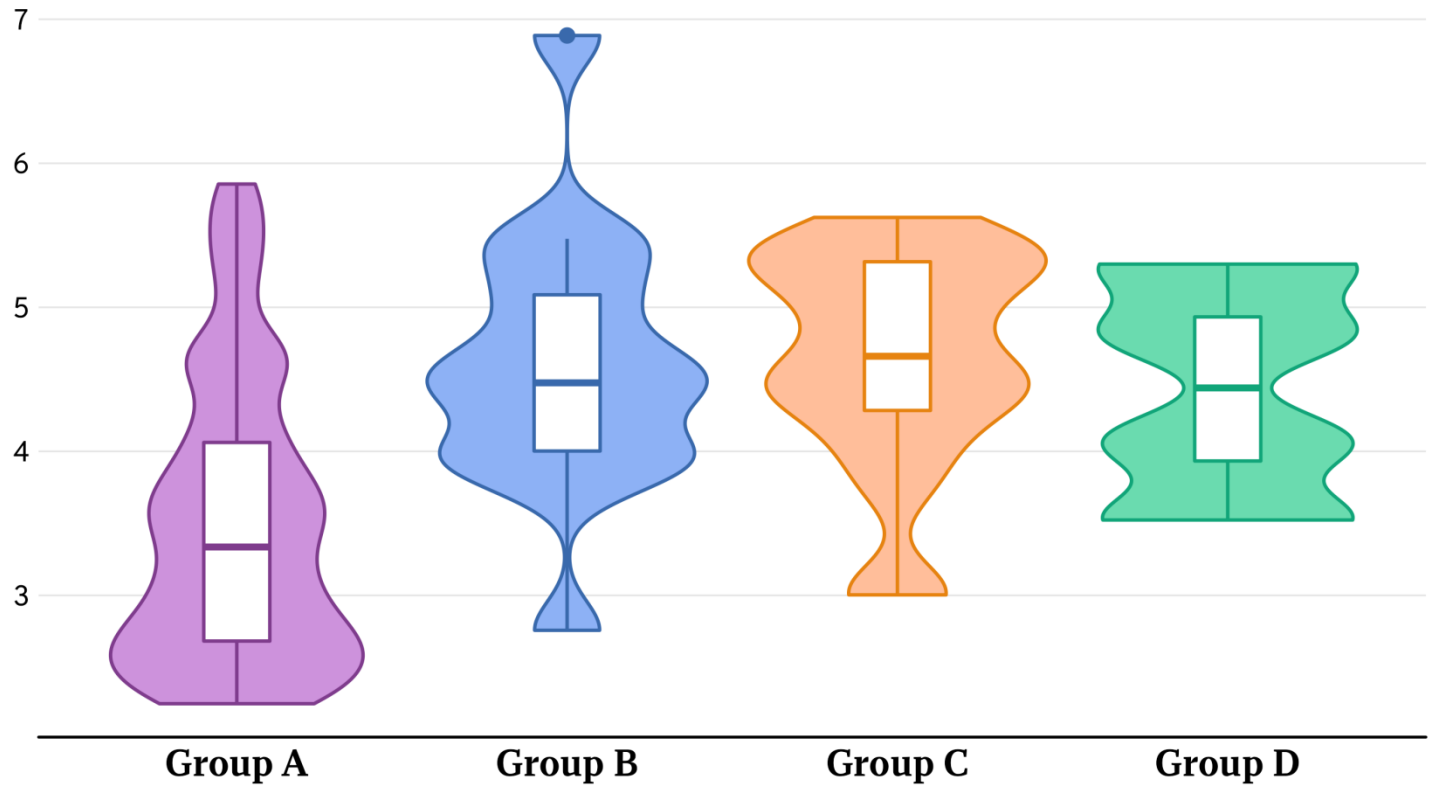
BOX PLOT X JITTER STRIPS

geom_boxplot() + geom_point(position = position_jitter())



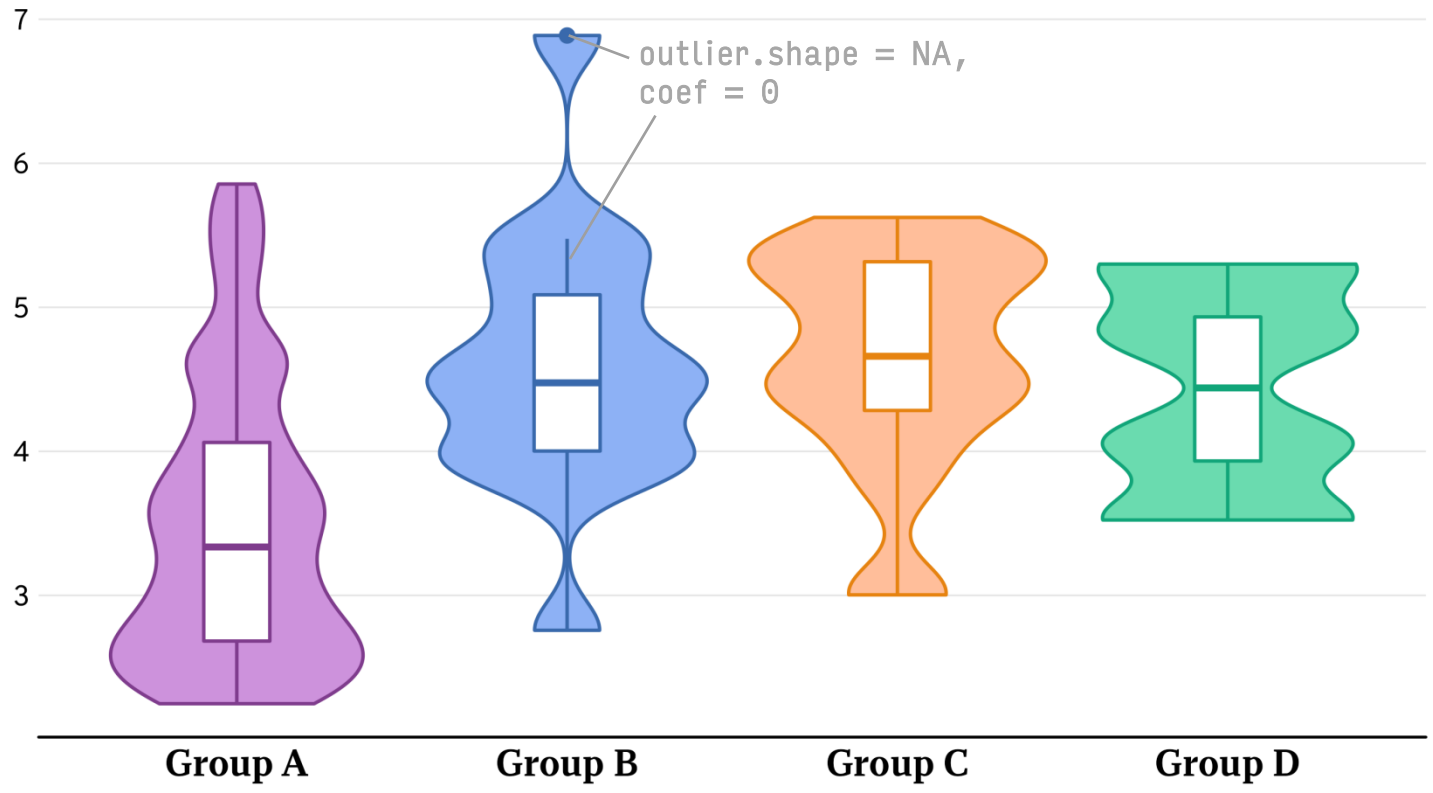
VIOLIN PLOT X BOX PLOT

geom_violin() + geom_boxplot()



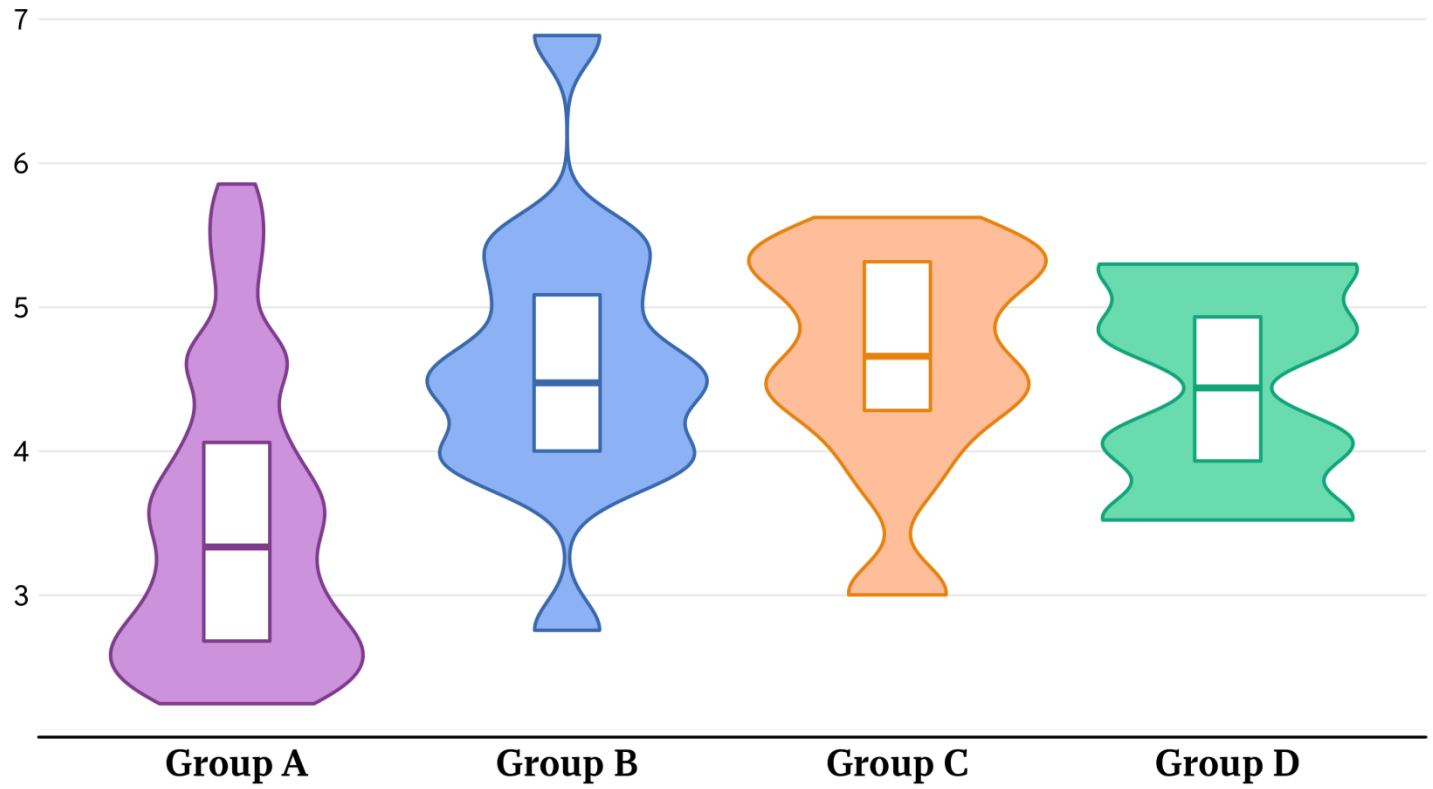
VİOLİN PLOT X BOX PLOT

geom_violin() + geom_boxplot()



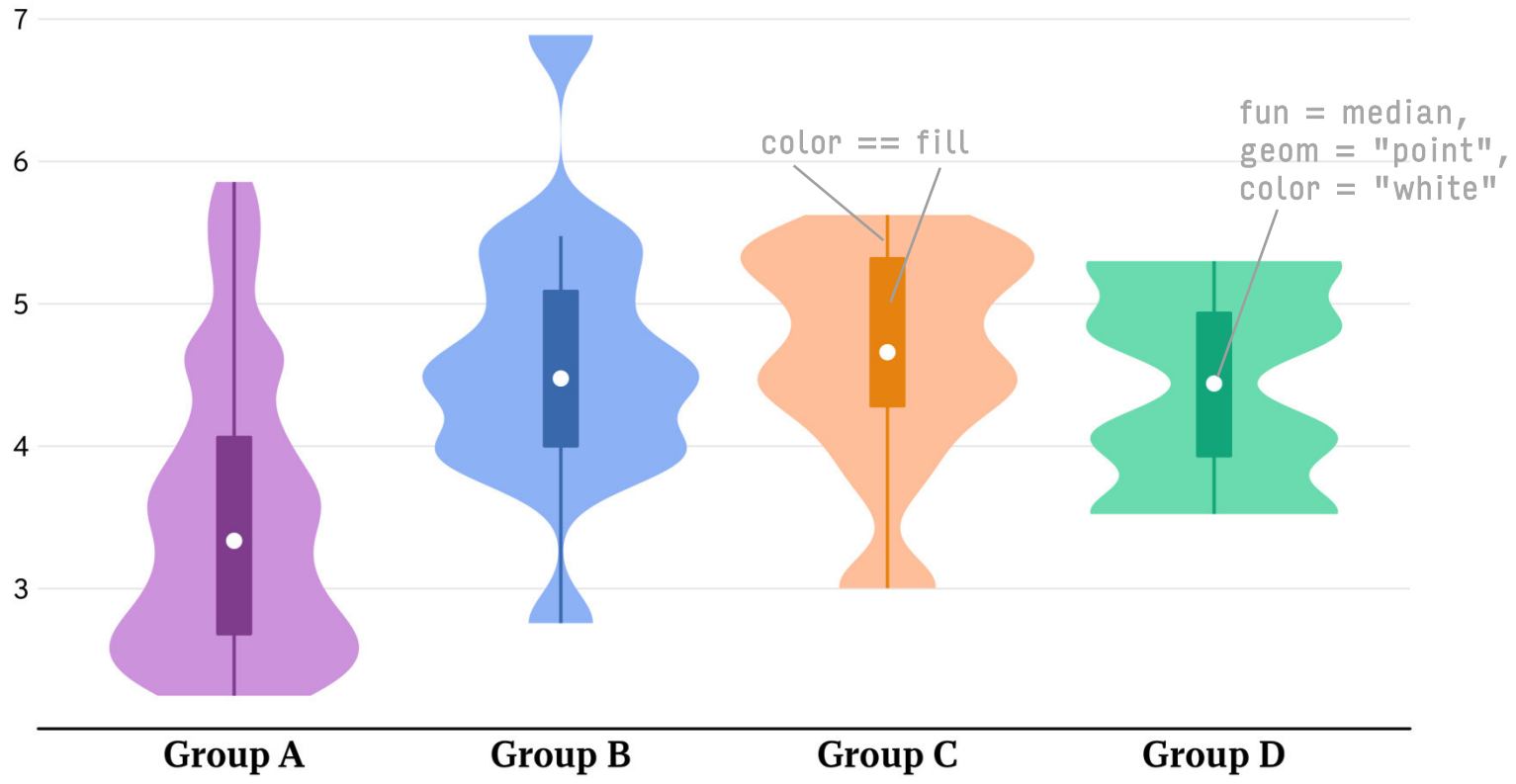
VIOLIN PLOT X BOX PLOT

geom_violin() + geom_boxplot()



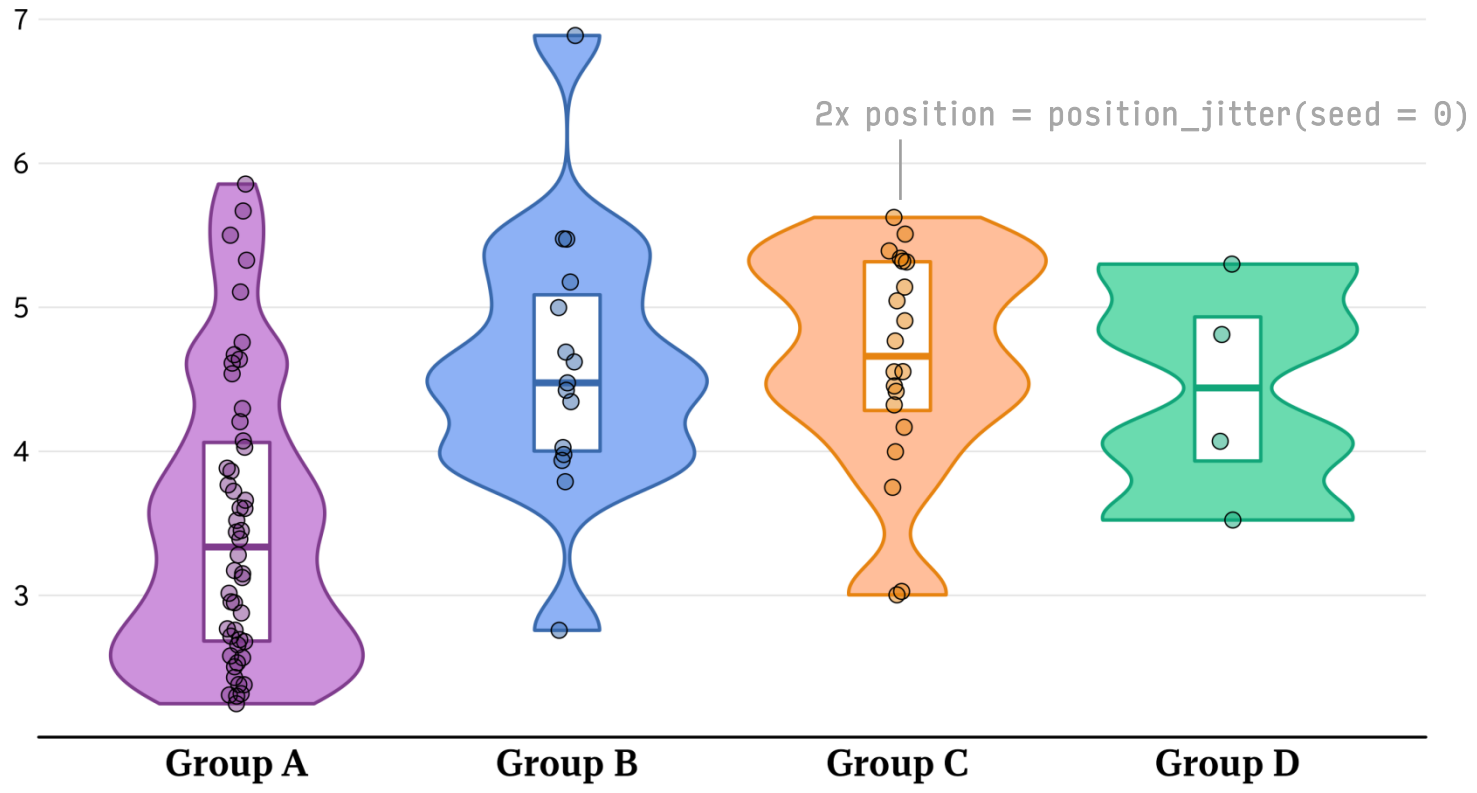
VIOLIN PLOT X BOX PLOT

```
geom_violin() + geom_boxplot() + stat_summary(geom = "point")
```



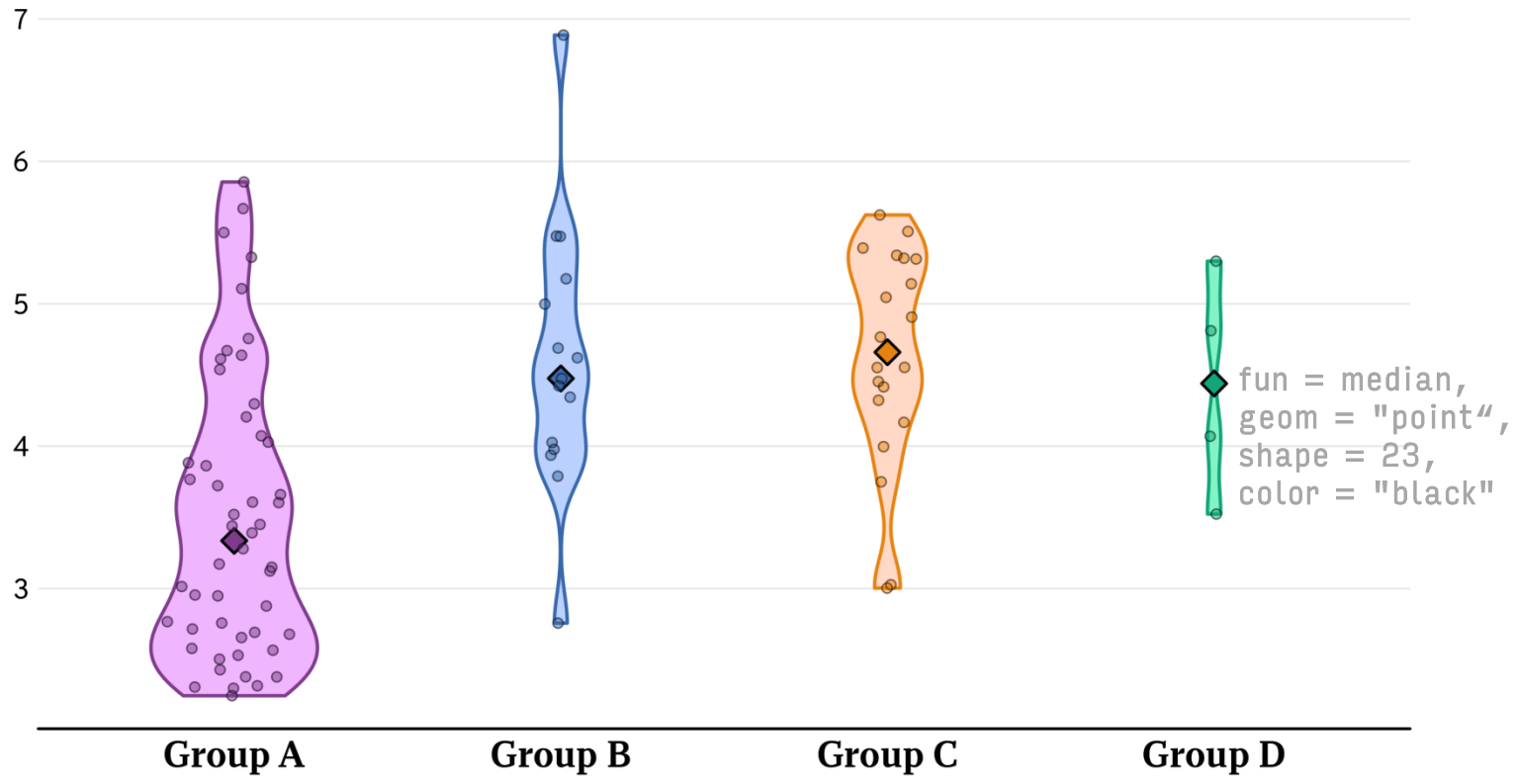
VİOLİN PLOT X BOX PLOT X JİTTER STRİPS

geom_violin() + geom_boxplot() + geom_point()



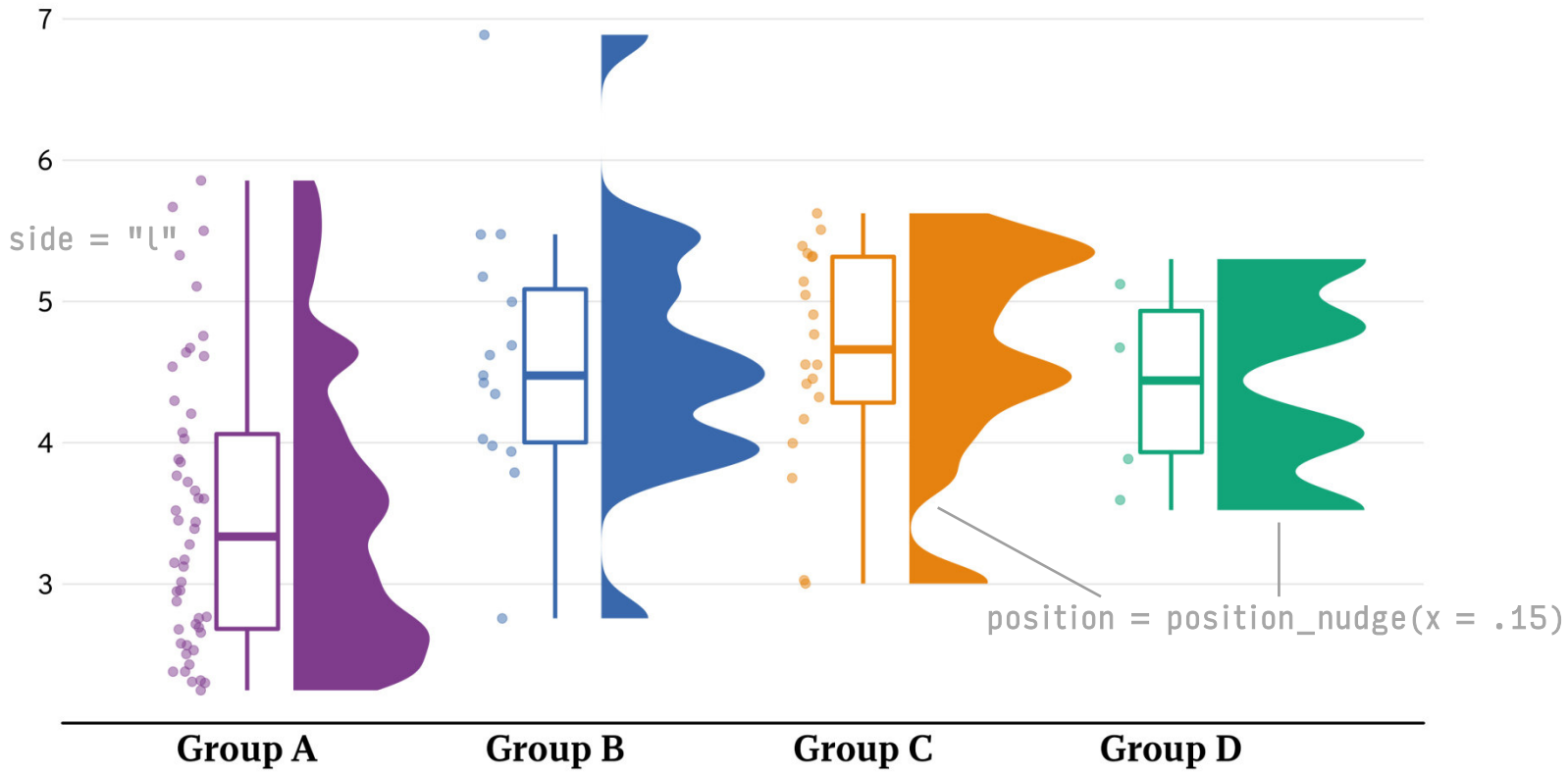
VİOLİN PLOT X SİNA PLOT

geom_violin(scale = "count") + ggforce::geom_sina() + stat_summary()



RAINCLOUD PLOTS

gghalves::geom_half_point() + geom_boxplot() + ggdist::stat_halfeye()



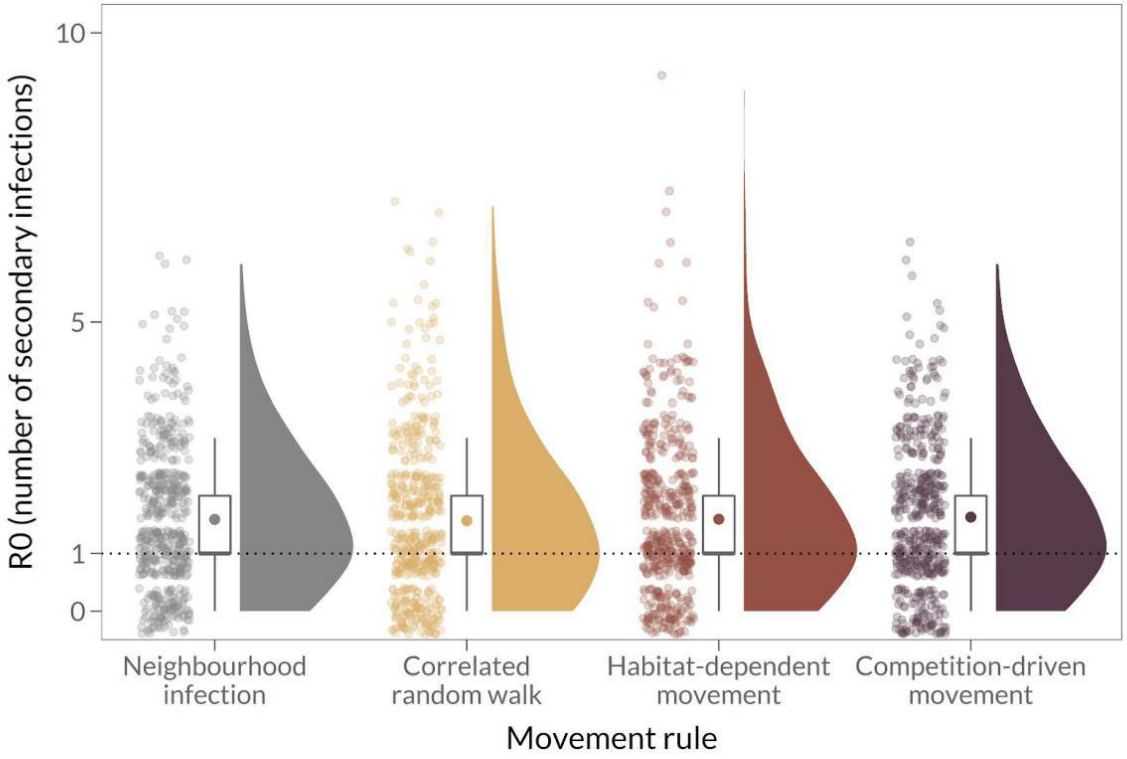
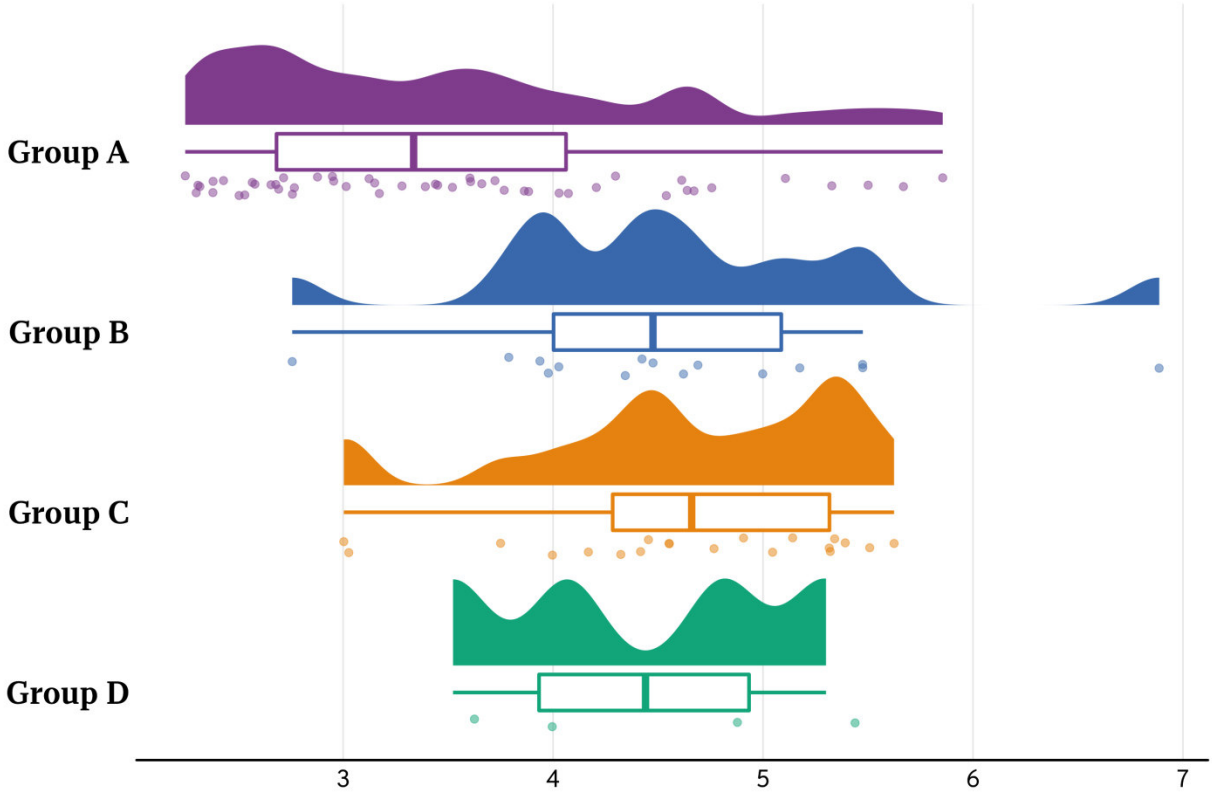


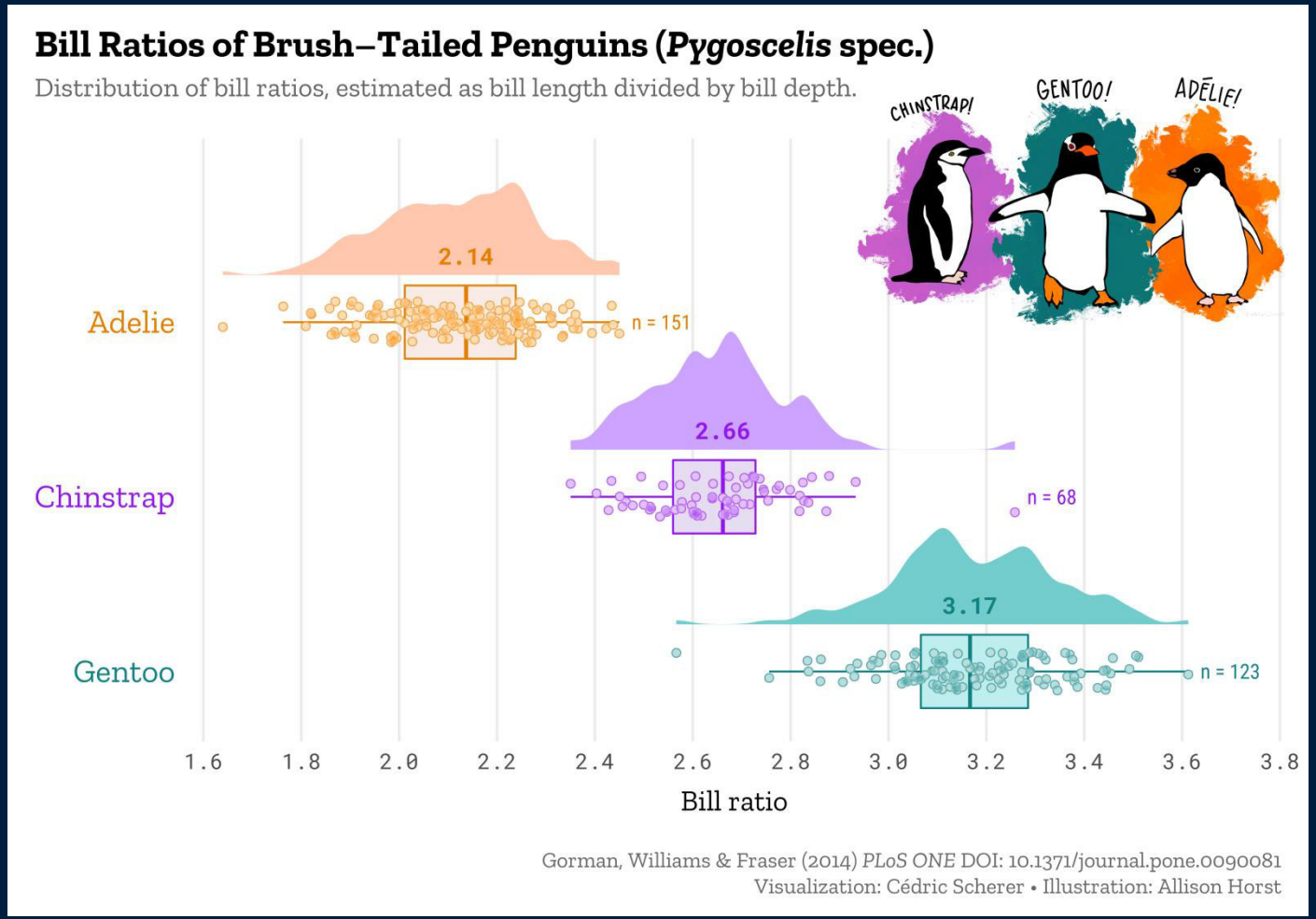
Figure S5: Probability distributions of the basic reproduction number R_0 for the on-the-move transmission probability using the best fit (β_m = 0.023; see Fig. S4) for each movement rule compared to the non-movement model using neighbourhood infection. The dotted line indicates an R_0 threshold value of 1.

Scherer et al. (2020) Oikos 129 (5):651–667 (10.1111/oik.07002)

RAINCLOUD PLOTS

`gghalves::geom_half_point() + geom_boxplot() + ggdist::stat_halfeye()`

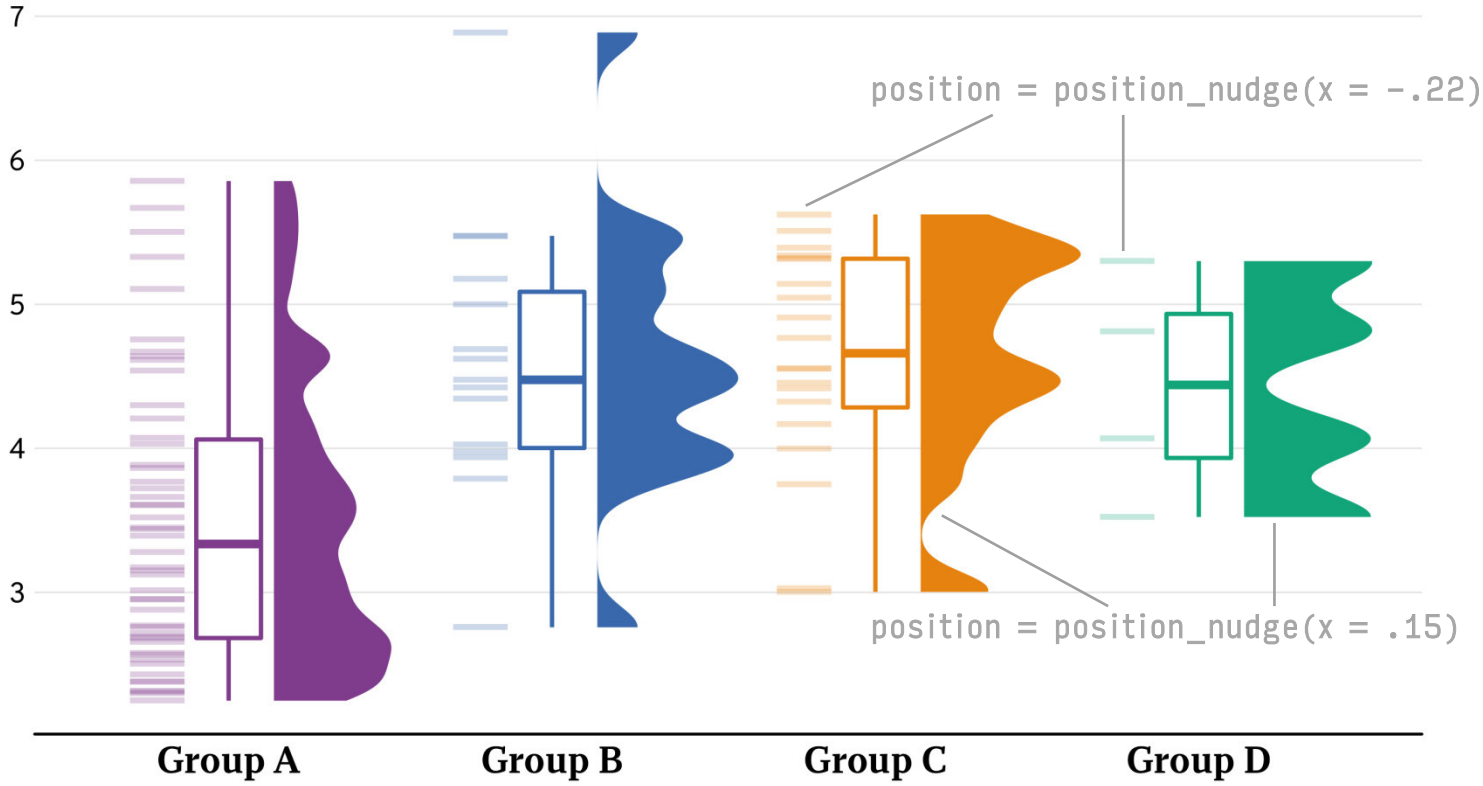




"Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)"

RAINCLOUD PLOTS WITH BARCODE STRIPS

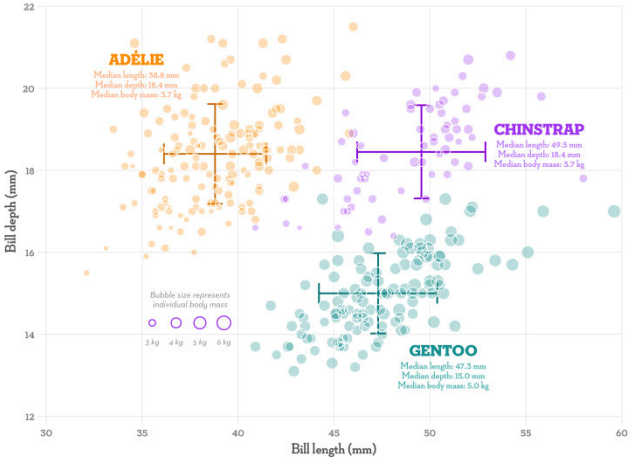
geom_point(shape = 95) + geom_boxplot() + ggdist::stat_halfeye()



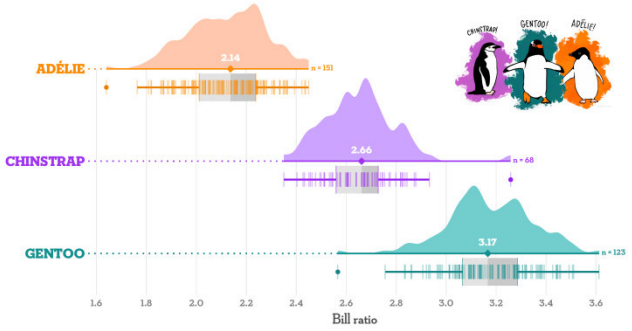
BILL DIMENSIONS OF BRUSH-TAILED PENGUINS

Pygoscelis adéliae (Adélie penguin) • P. antarctica (Chinstrap penguin) • P. papua (Gentoo penguin)

A. Scatterplot of bill length versus bill depth (median +/- sd)

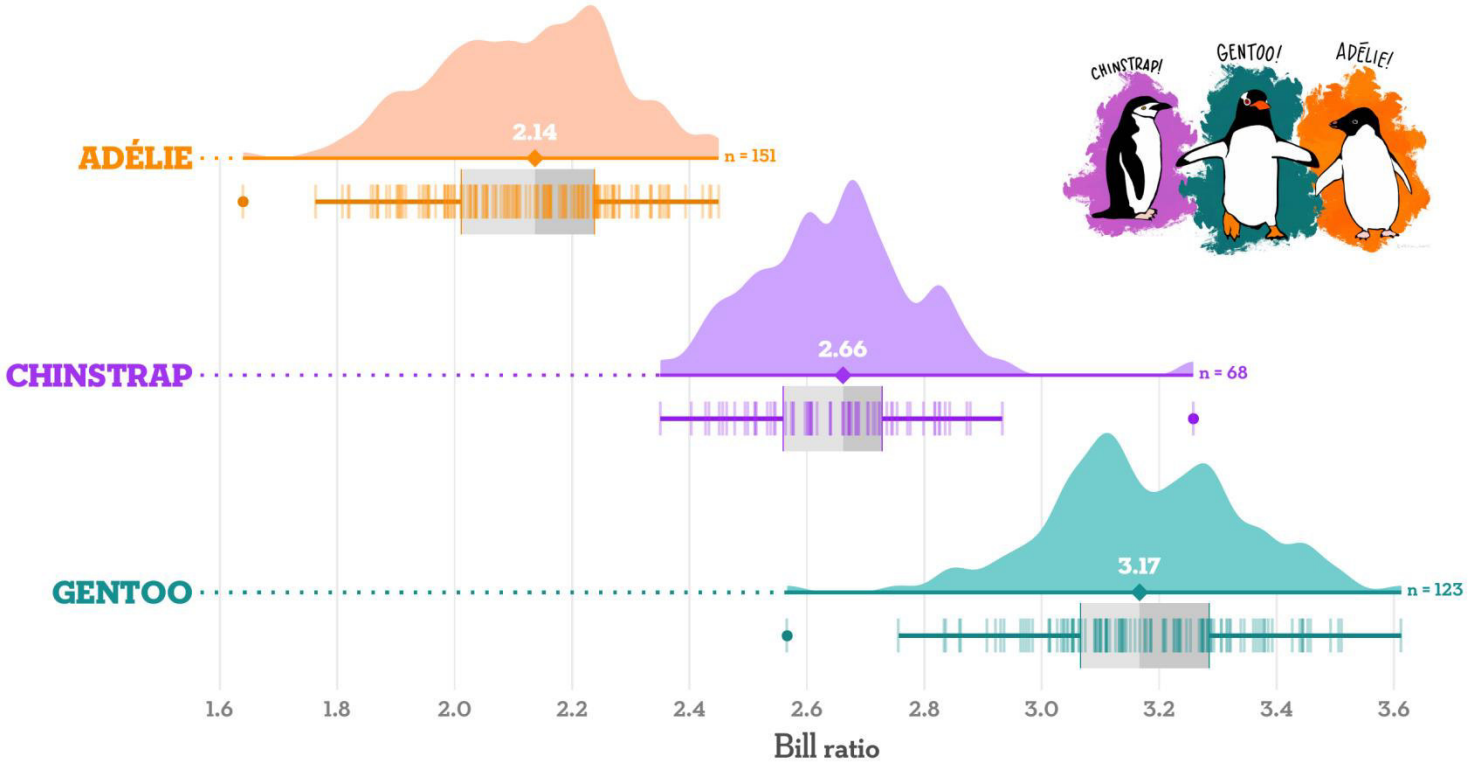


B. Distribution of the bill ratio, estimated as bill length divided by bill depth



Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0090081 • Illustrations: Allison Horst

B. Distribution of the bill ratio, estimated as bill length divided by bill depth

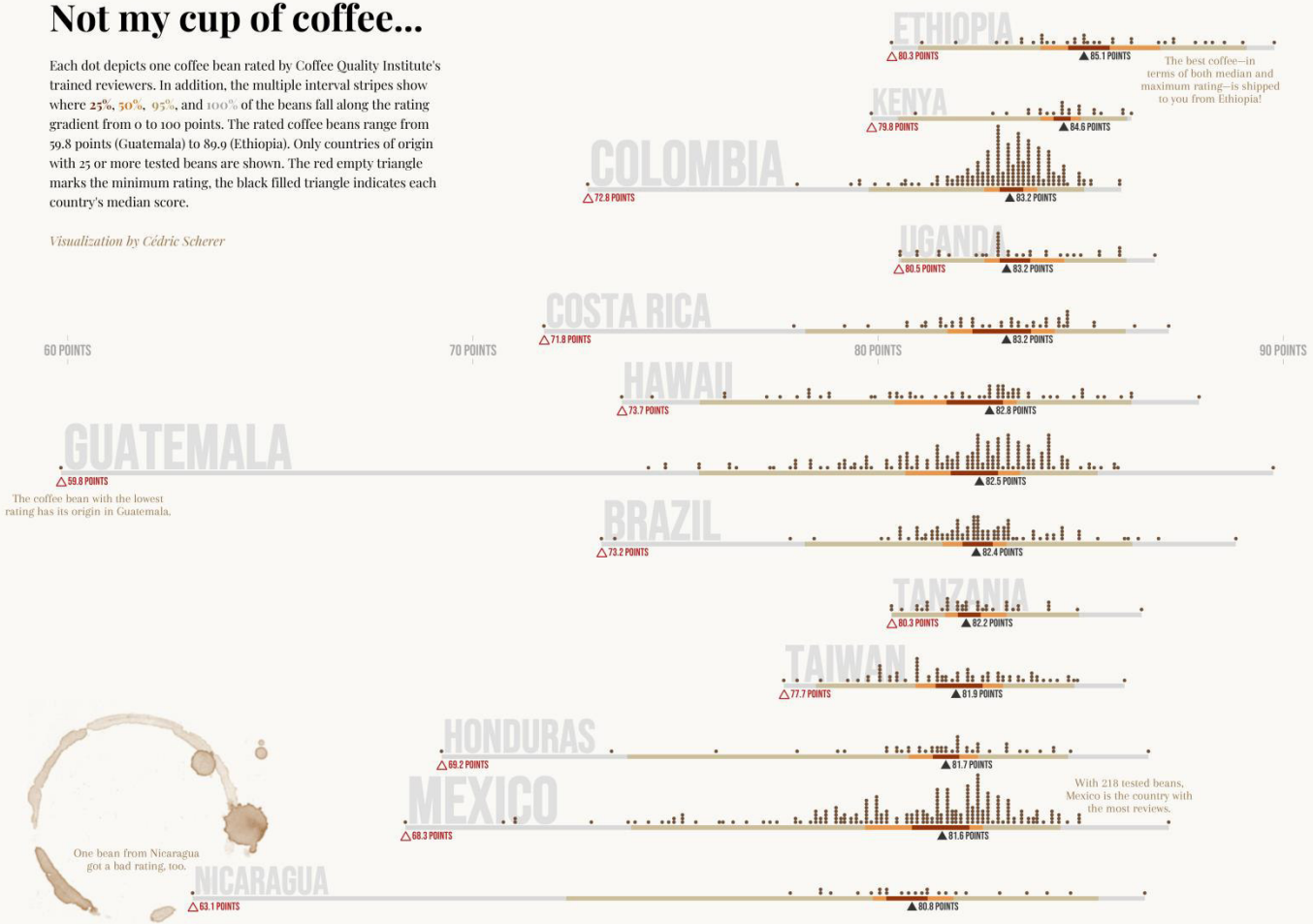


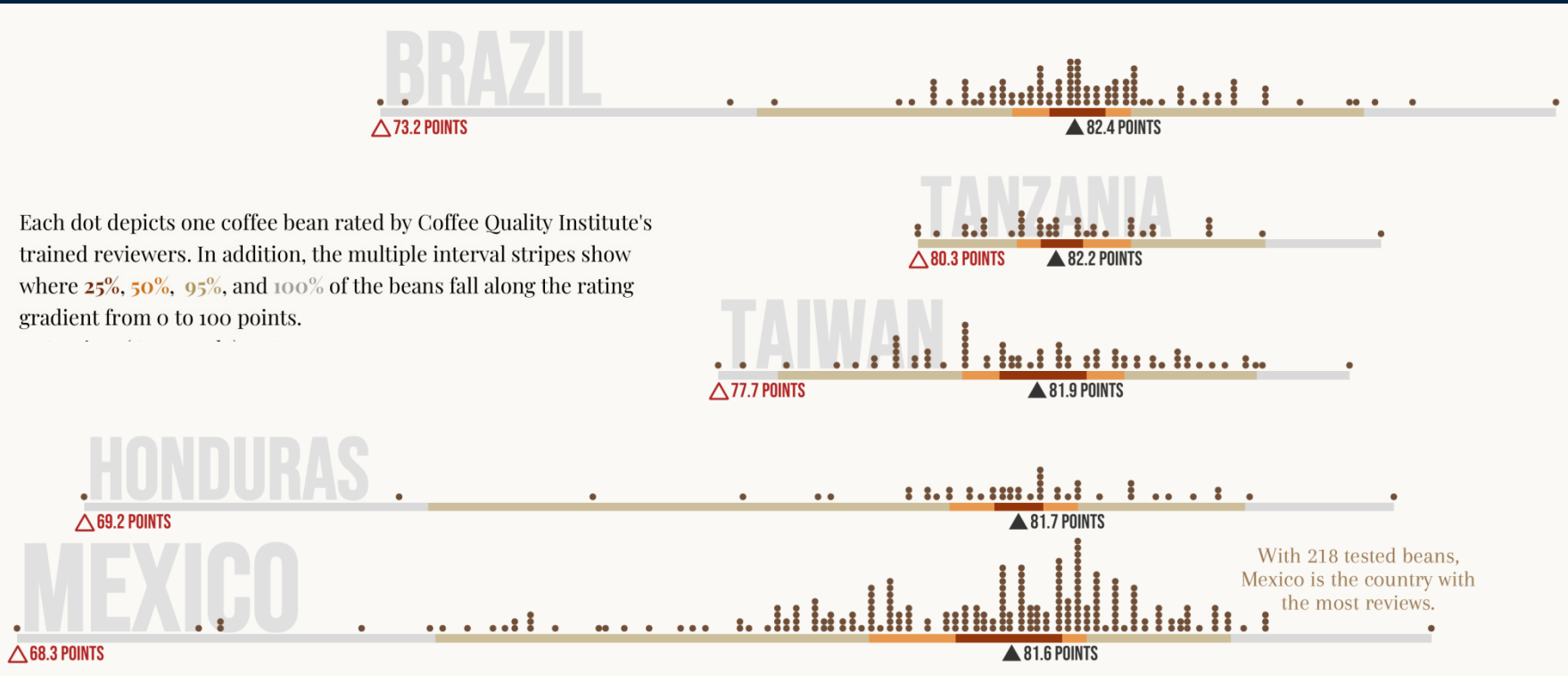
Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0090081 • Illustrations: Allison Horst

Not my cup of coffee...

Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points. The rated coffee beans range from 59.8 points (Guatemala) to 89.9 (Ethiopia). Only countries of origin with 25 or more tested beans are shown. The red empty triangle marks the minimum rating, the black filled triangle indicates each country's median score.

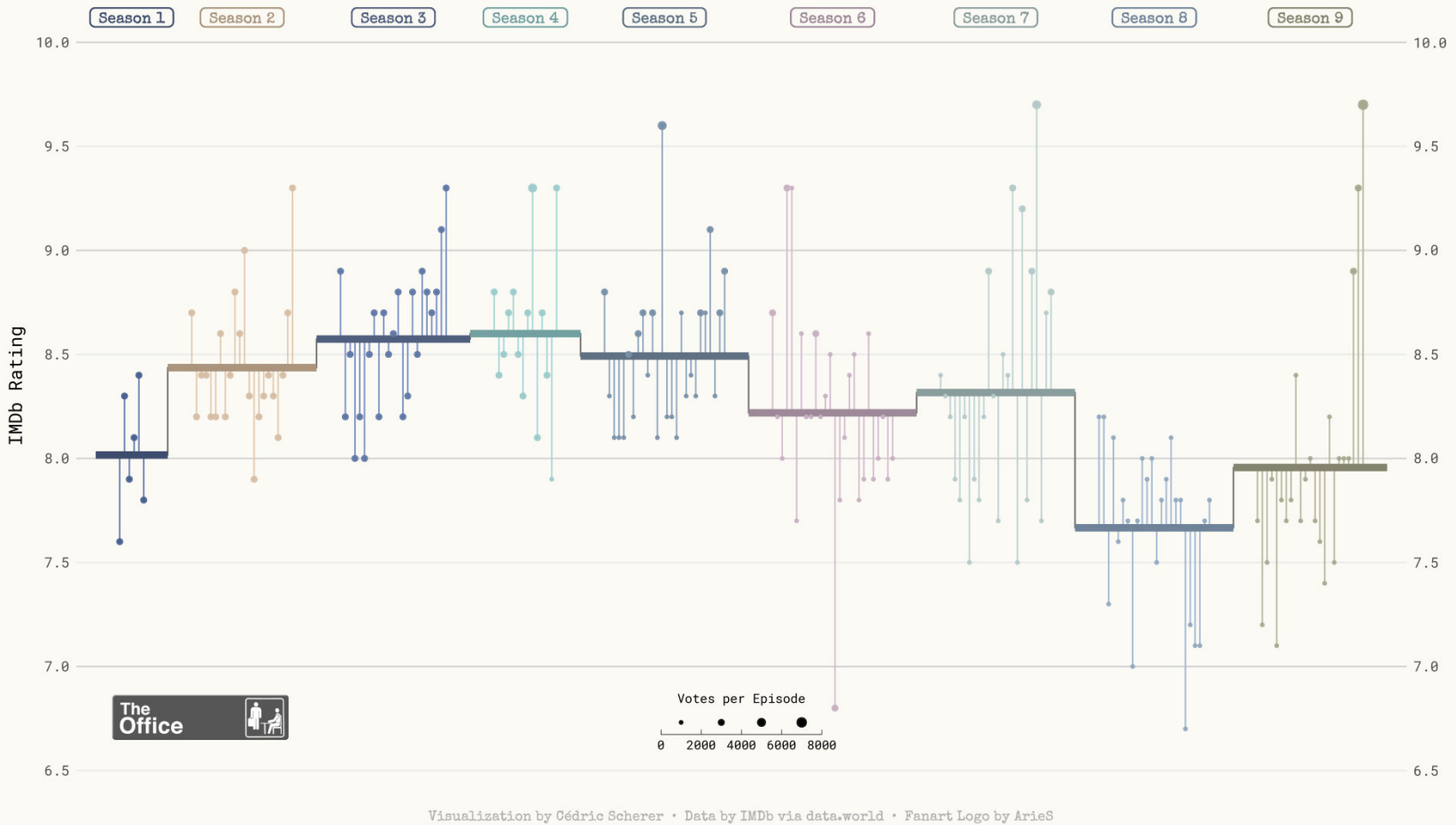
Visualization by Cédric Scherer



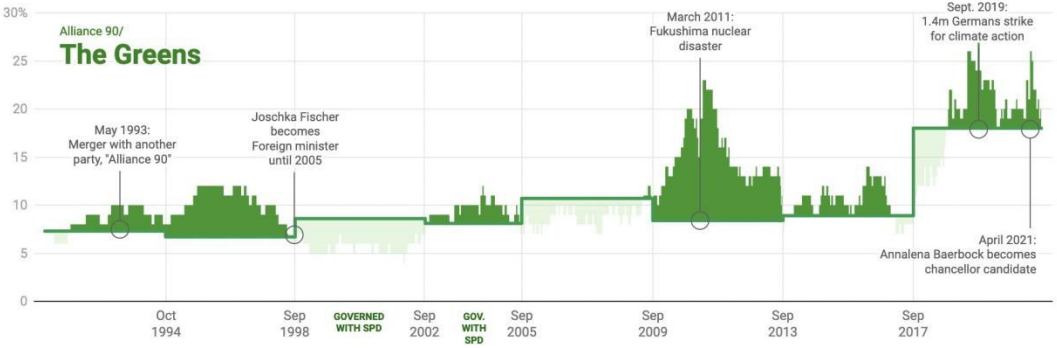

BRAZIL
73.2 POINTS
82.4 POINTS
TANZANIA
80.3 POINTS
82.2 POINTS
TAIWAN
77.7 POINTS
81.9 POINTS
HONDURAS
69.2 POINTS
81.7 POINTS
MEXICO
68.3 POINTS
81.6 POINTS
Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points.
With 218 tested beans, Mexico is the country with the most reviews.

SOMETHING

Special

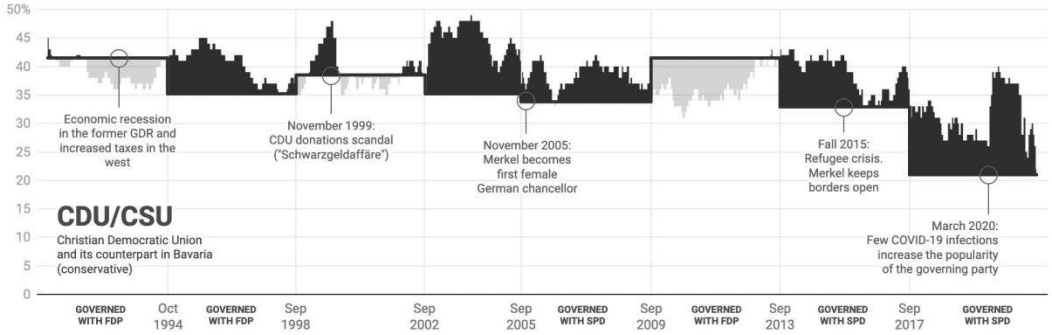


Visualization by Cédric Scherer • Data by IMDb via data.world • Fanart Logo by ArieS



Polls compared with the results of the next federal election. The straight line going from 2017 to 2021 is based on polls a month before the election.

Chart: Lisa Charlotte Rost, Datawrapper • Source: Forschungsgruppe Wahlen: Politbarometer • Get the data



Polls compared with the results of the next federal election. The straight line going from 2017 to 2021 is based on polls a month before the election.

Chart: Lisa Charlotte Rost, Datawrapper • Source: Forschungsgruppe Wahlen: Politbarometer • Get the data

"Which German party is the most unlucky when it comes to election dates?" *by Lisa C. Muth (DataWrapper)*

"ANIMAL CROSSING" SEEMS TO BE FUN, NO MATTER IF YOU LIKE IT OR NOT

Or, more likely, users rating it low wrote things like "not much fun", "no fun at all" or "A fun game that I really wanted to play but what a **** multiplayer mode."

Low Grades (0–2)

Medium Grades (3–7)

High Grades (8–10)

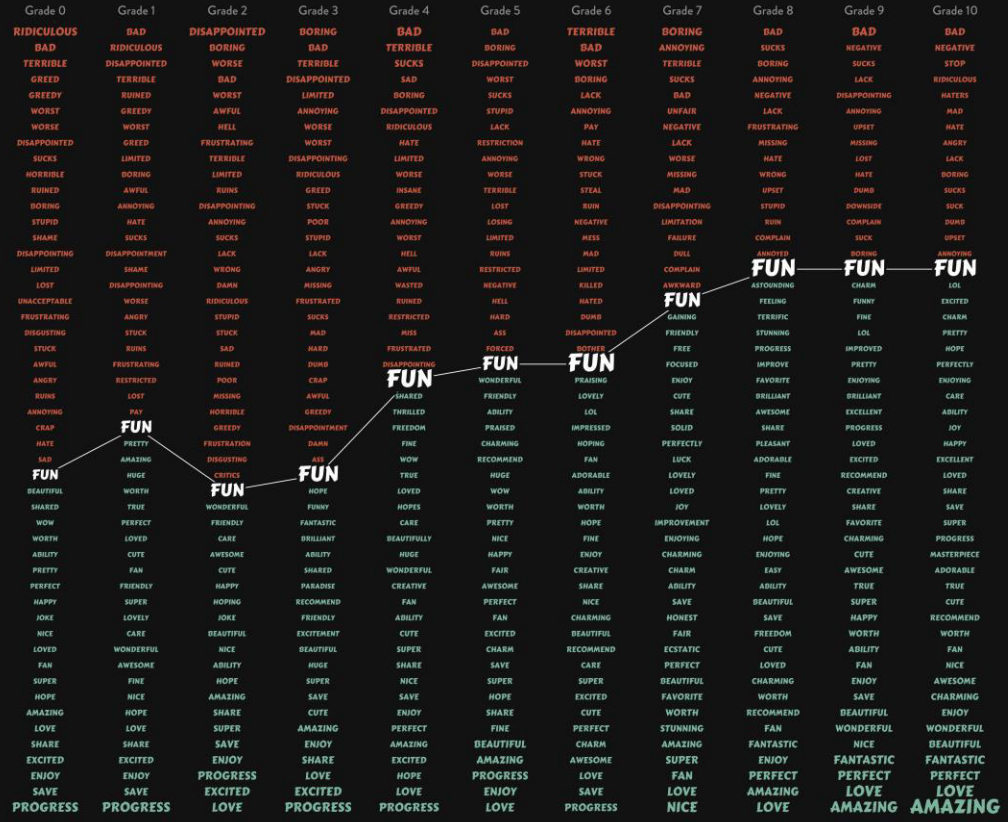
Based on a sentiment analysis of "Animal Crossing" user reviews, each wordcloud shows the 50 words that contributed the most to each grade category, either in a positive or in a negative way.

Visualization by Cédric Scherer • Data by Metacritic

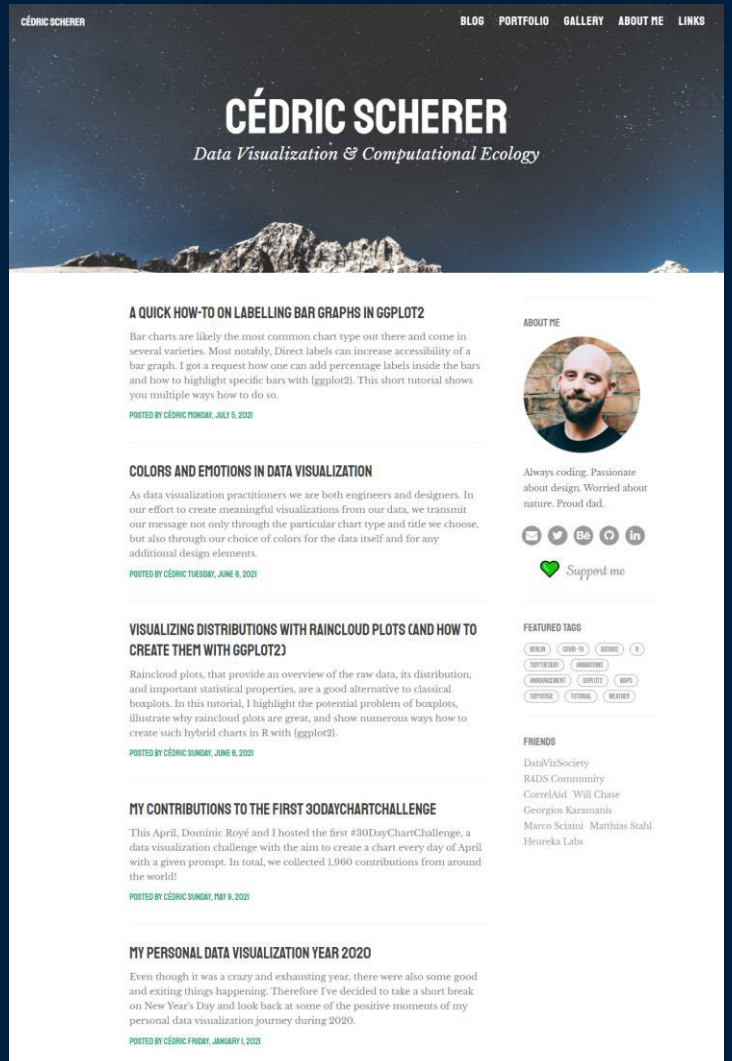
"ANIMAL CROSSING" SEEMS TO BE FUN, NO MATTER IF YOU LIKE IT OR NOT

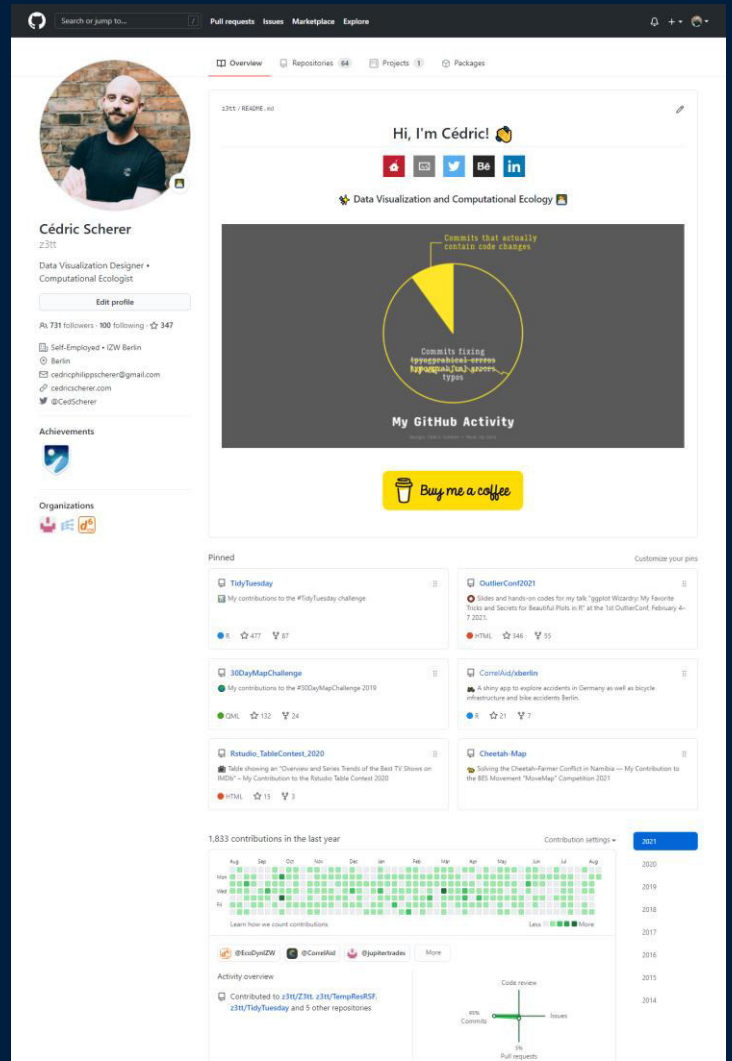
Or, more likely, users rating it low wrote things like "not much fun", "no fun at all" or "A fun game that I really wanted to play but what a **** multiplayer mode.

Based on a sentiment analysis of user reviews, each "word stripe" shows the 50 words that contributed the most to each grade category, either in a positive or in a negative way. The size of each word indicates its contribution per grade. The amount of words associated with positive sentiment increase, as we would expect, with higher grades. Fun, even though categorized as positive, is often used in both contexts.



Visualization by Cédric Scherer • Data by Metacritic





THANK YOU!

cedricscherer.com

twitter.com/CedScherer

github.com/z3tt

Buy me a coffee

buymeacoffee.com/z3tt