

INTRODUCTION TO GGPLOT2



Welcome!

Schedule

1. Introduction to **{ggplot2}**

2. **{ggplot2}** concepts

3. Advanced customization

4. Extensions

Prerequisites

You need a laptop with the following installed:

- R version >= 4.1 and RStudio >= 2022.02.1
- R package development toolchain: Rtools(windows) or XCode (mac) or r-base-dev
- The following packages:

```
my_packages <- c("rmarkdown", "knitr", "tidyverse", 'patchwork',
                 "cowplot", 'colorspace', 'ggtext', 'ggokabeito',
                 'ggthemes', 'ggdist', 'GGally', 'ggrepel', 'ggpubr',
                 'ggstatsplot', 'ggdag', 'plotly', 'ggiraph')
install.packages(my_packages)
install.packages("devtools")
devtools::install_github("haleyjeppson/NCME23data")
devtools::install_github("kassambara/ggcorrplot")
```

WiFi

- Username: AERA2023
- Password: AERA2023

Outline

1. Set up
2. Intro to `{ggplot2}`
3. Motivating example

Get started

1. Get workshop materials

- Clone the [workshop repo](#) from GitHub to work locally
- Follow along [here](#)

2. Create a new R Markdown document or R script

- You can follow along in either an .Rmd or R script

3. Attach R packages

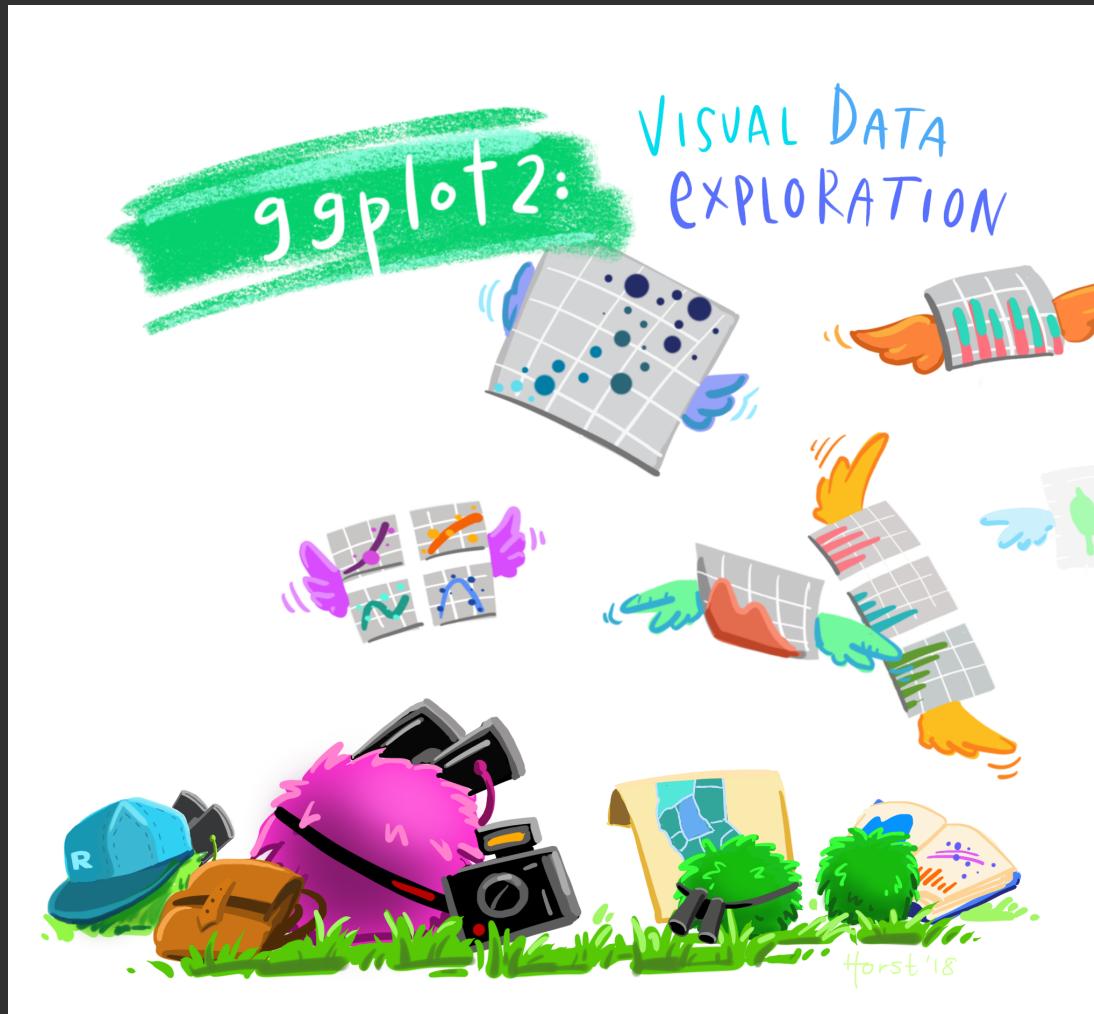
```
# General use packages:  
library(tidyverse)  
library(NCME23data)
```

Learning Goals

Upon completion of this workshop, you will be able to:

1. **identify** the appropriate plot types and corresponding `{ggplot2}` geoms to consider when visualizing your data;
2. **implement** the ggplot2 grammar of graphics by using `ggplot()` and build up plots with the `+` operator;
3. **iterate** through multiple visualizations of your data by changing the aesthetic mappings, geometries, and other graph properties;
4. **incorporate** interaction and custom elements (colors, fonts, etc.) into your visualizations by adjusting `{ggplot2}` theme elements; and
5. **investigate** the world of `{ggplot2}` independently to expand upon the skills learned in the course.

Introduction to {ggplot2}



Artwork by @allison_horst

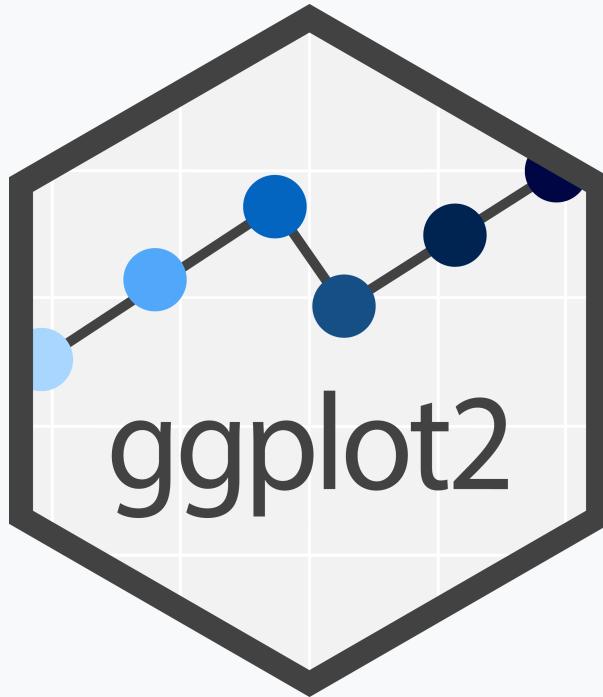
The `{ggplot2}` Package

... is an R package to visualize data created by Hadley Wickham in 2005

```
# install.packages("ggplot2")
library(ggplot2)
```

... is part of the `{tidyverse}`

```
# install.packages("tidyverse")
library(tidyverse)
```



{ggplot2} is a system for declaratively creating graphics, based on "The Grammar of Graphics"
(Wilkinson, 2005)

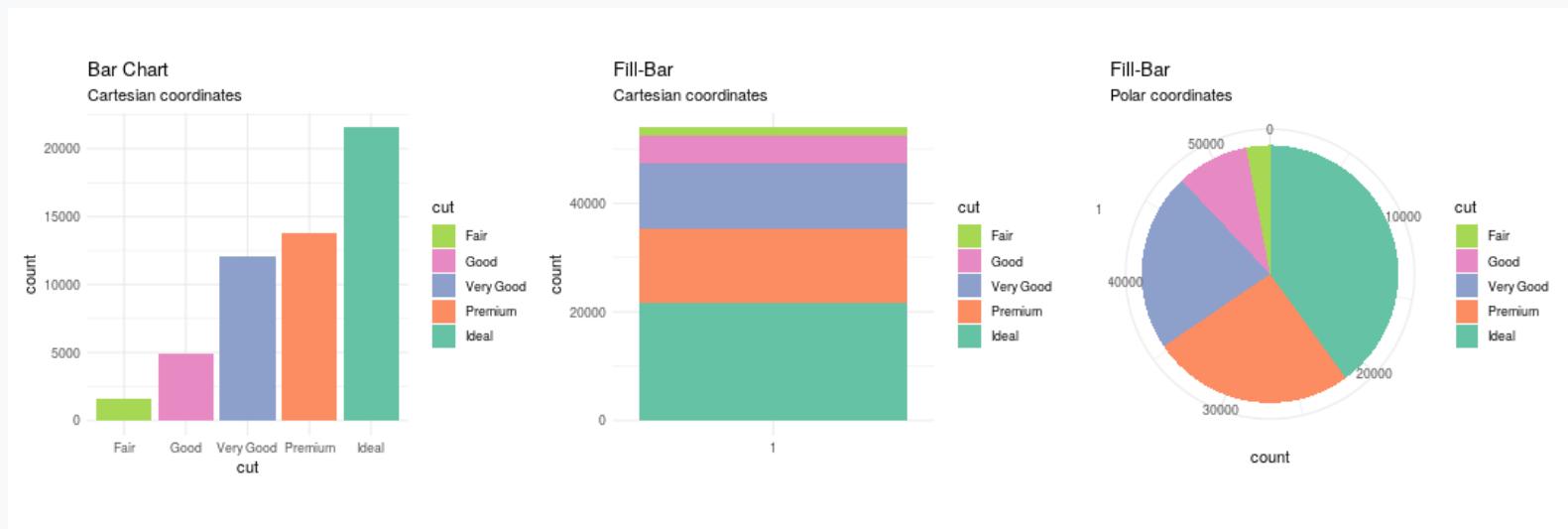
You provide the data, tell **{ggplot2}** how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

Why `{ggplot2}`?

- designed to adhere to good graphical practices (uses the grammar of graphics)
- flexible system that supports a wide variety of plot types
- constructs plots using the concept of layers
- lots of additional functionality thanks to extensions
- active and helpful community

Grammar of graphics

- Developed by Leland Wilkinson
- Think about graphical forms rather than a limited set of graphs
- Different types of graphs may appear completely distinct, but share many common elements.
- By making different visual choices, you can use graphs to highlight different aspects of the same data.

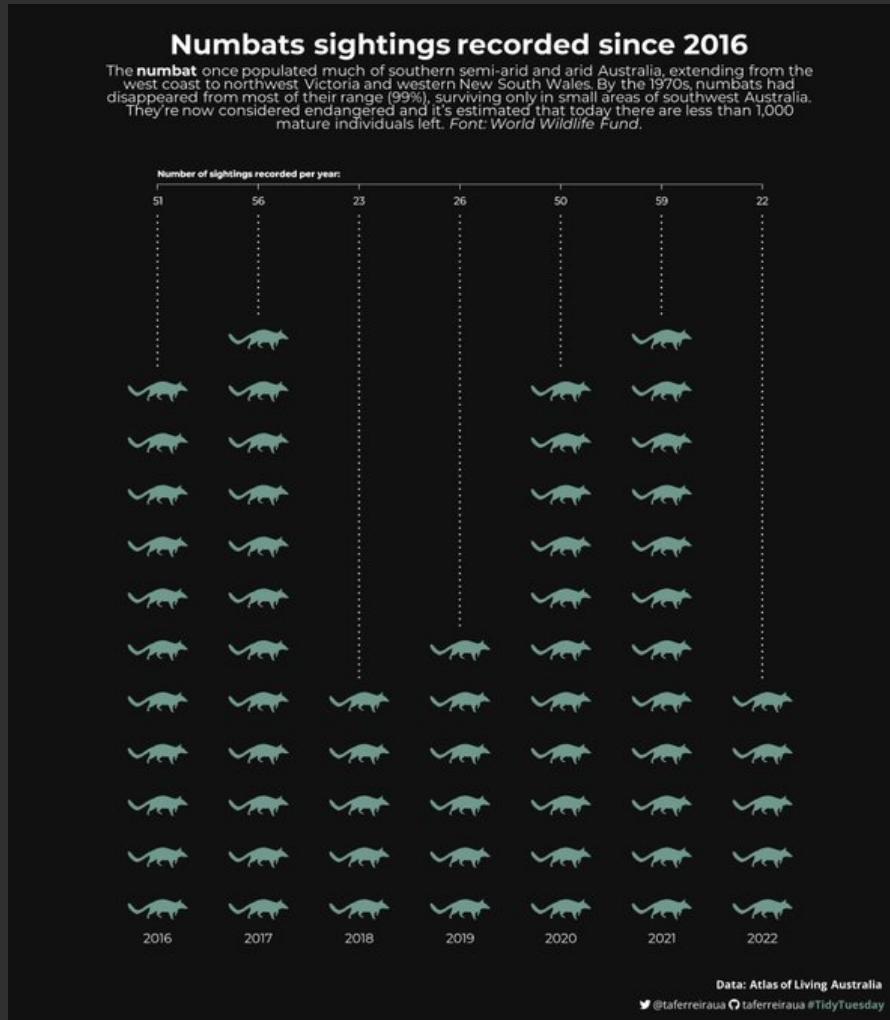


A layered grammar of graphics

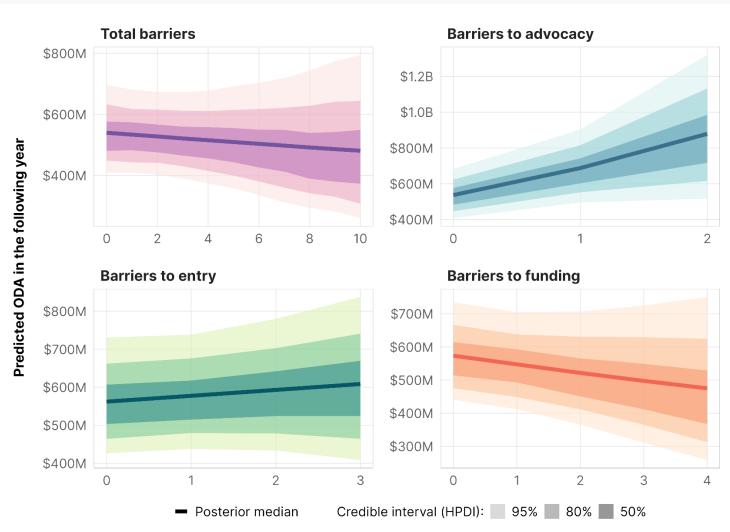
A graphical representation (plot) in ggplot2 consists of:

1. **default data and mappings (aes)**: data variables are mapped to visual properties of the graphical elements
2. **one or more layers**:
 - geometric element (**geom**, such as point, line, rectangle, text, ...),
 - statistical transformation (**stat**, such as identity, counts, bins, ...),
 - position adjustment,
 - (optional) one dataset and set of aesthetic mappings
3. **scales**: map values in the data space to values in an aesthetic space
4. **coordinate system (coord)**: normally Cartesian, but could use polar coordinates for pie charts or different mapping coordinates
5. **facetting**: for small multiples (subsets) and their arrangement
6. **theme**: fine-tune display items, such as font and its size, color of background, margins, ...

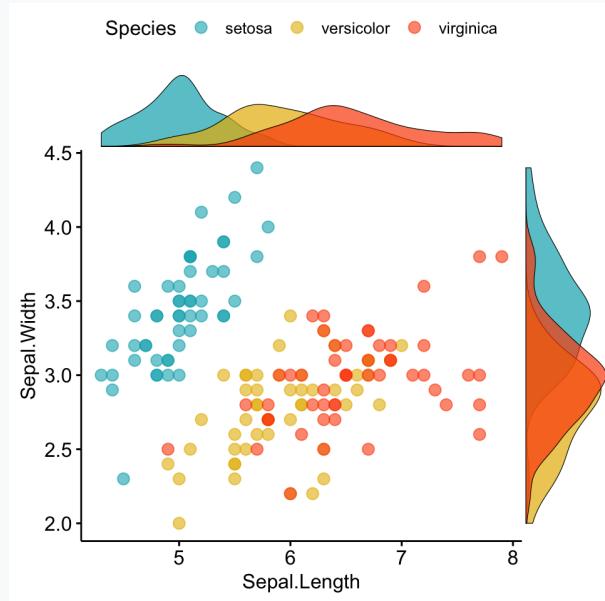
The `{ggplot2}` showcase



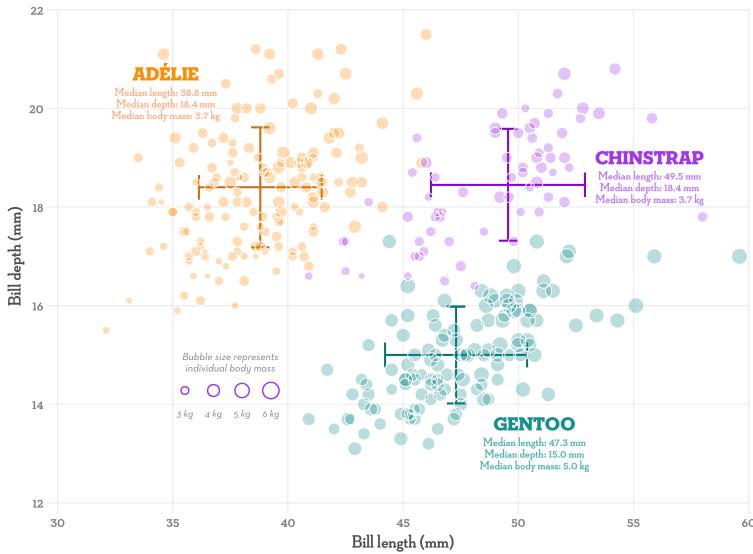
by Thays Ferreira



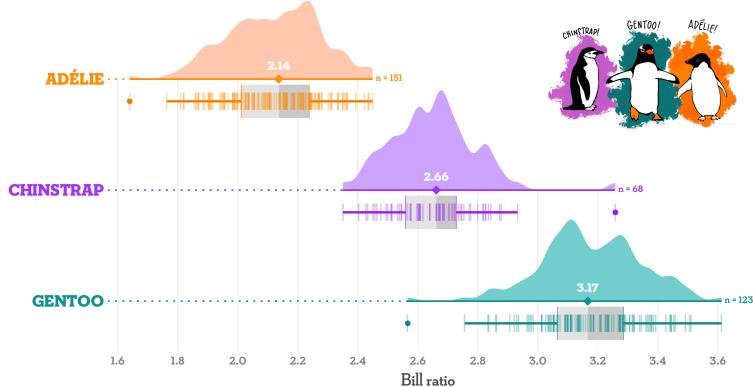
by Andrew Heiss



by Datanovia.com

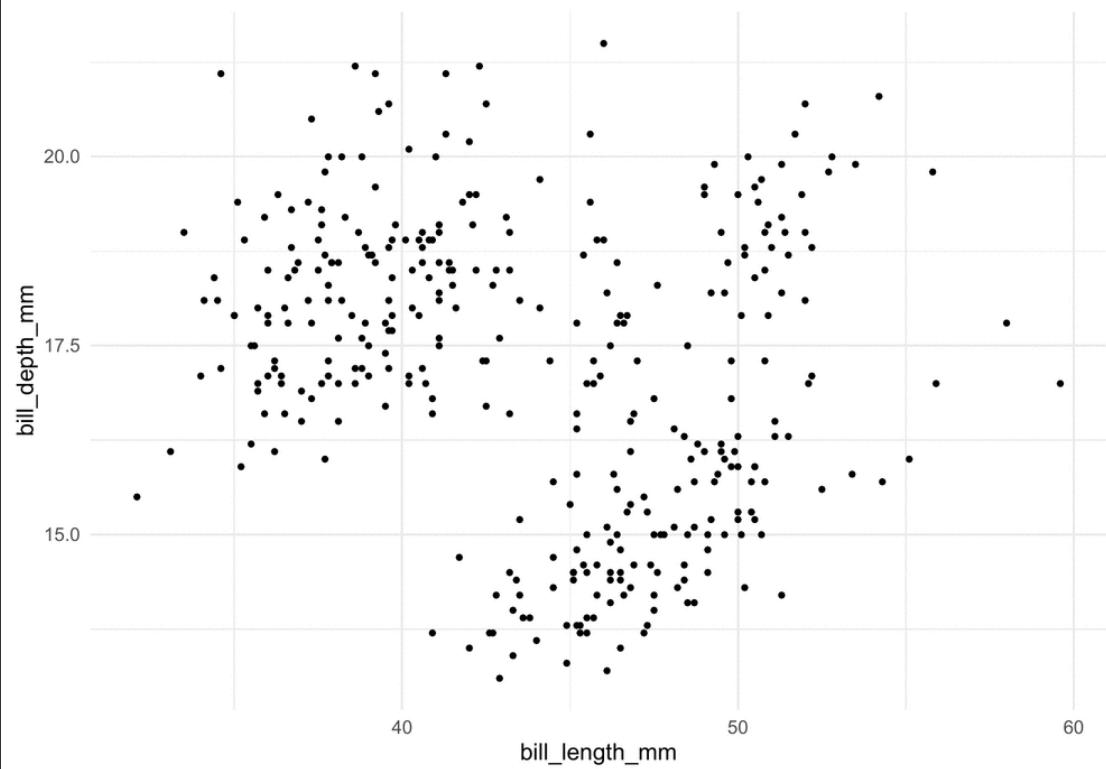


B. Distribution of the bill ratio, estimated as bill length divided by bill depth



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

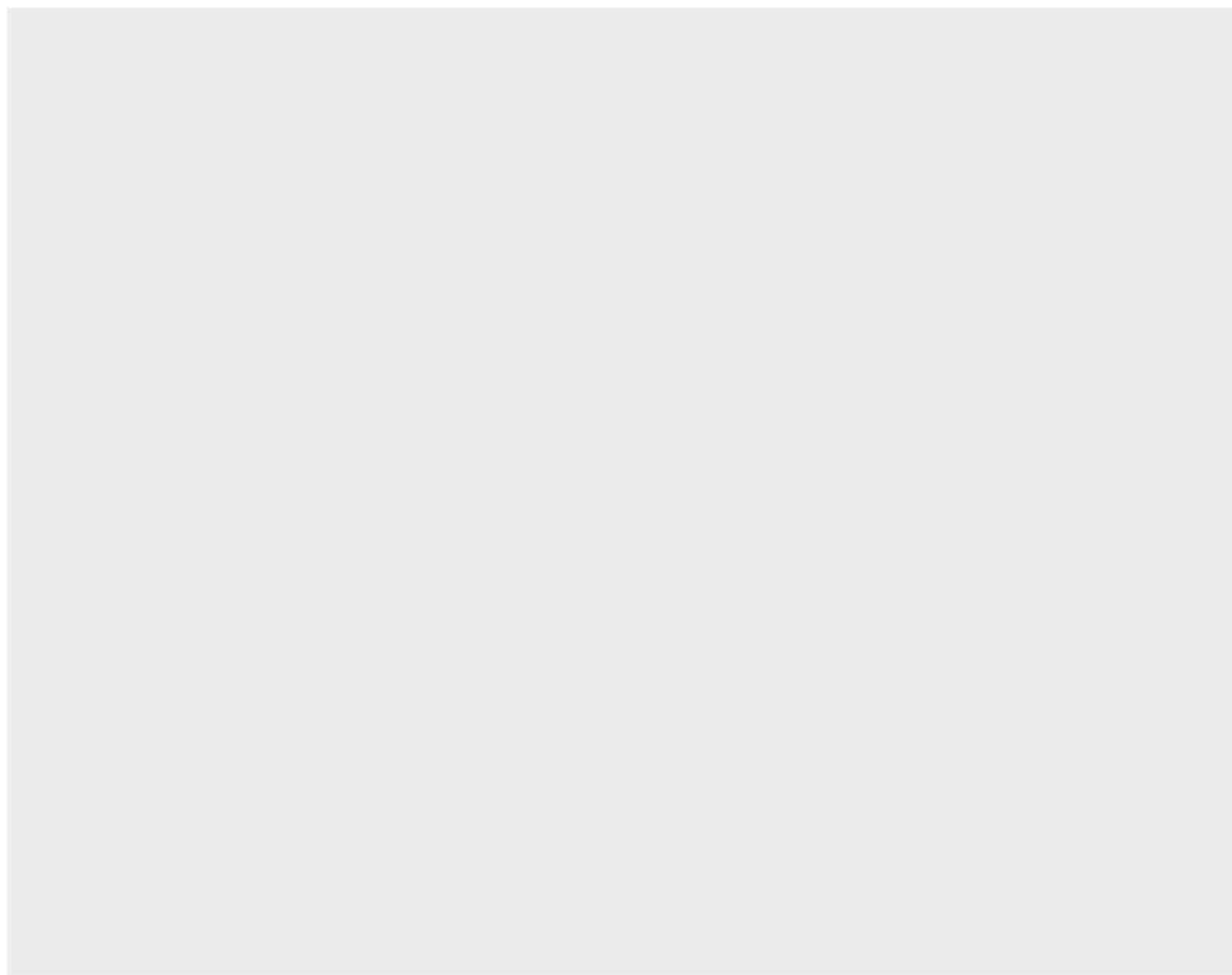
by Cédric Scherer

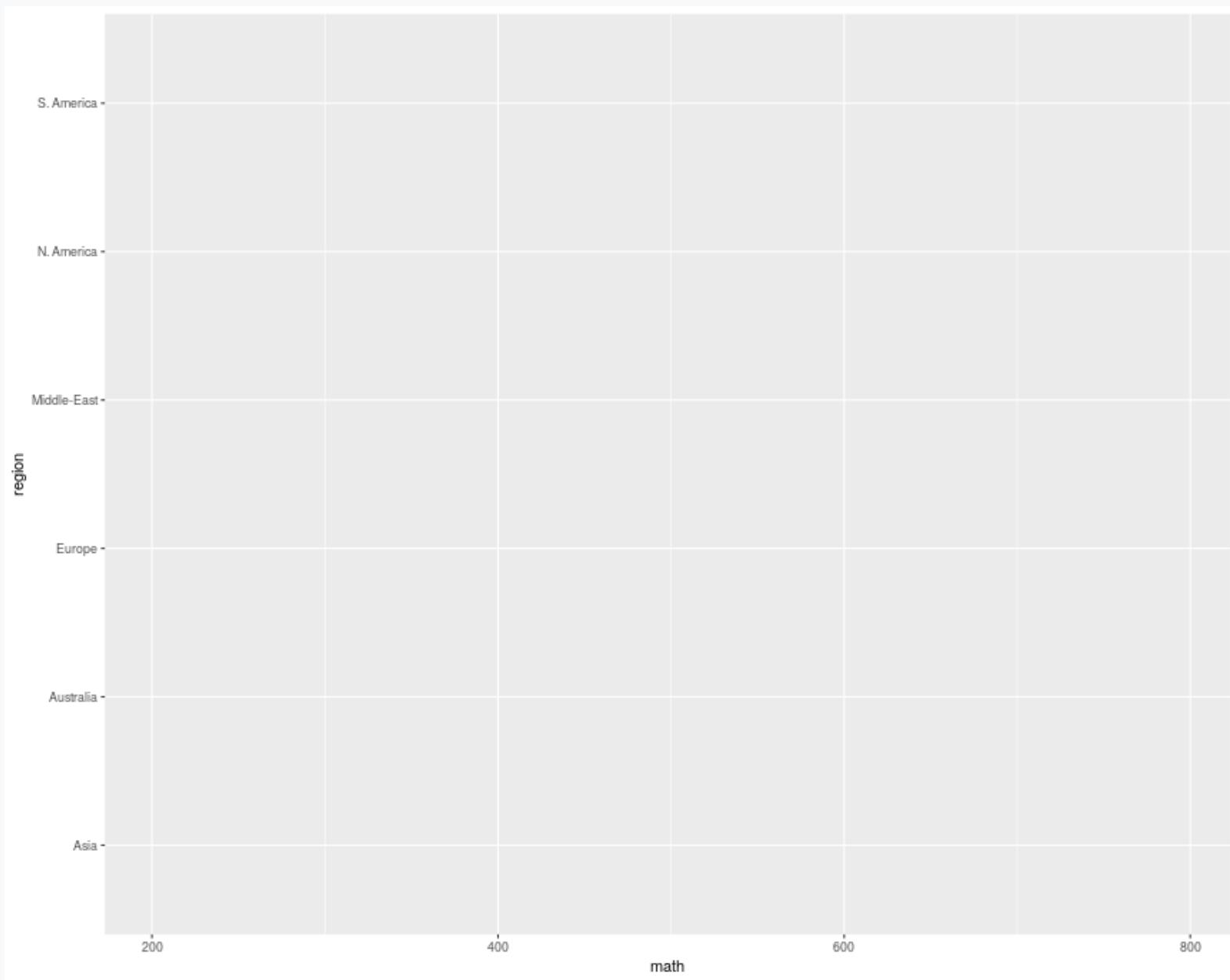


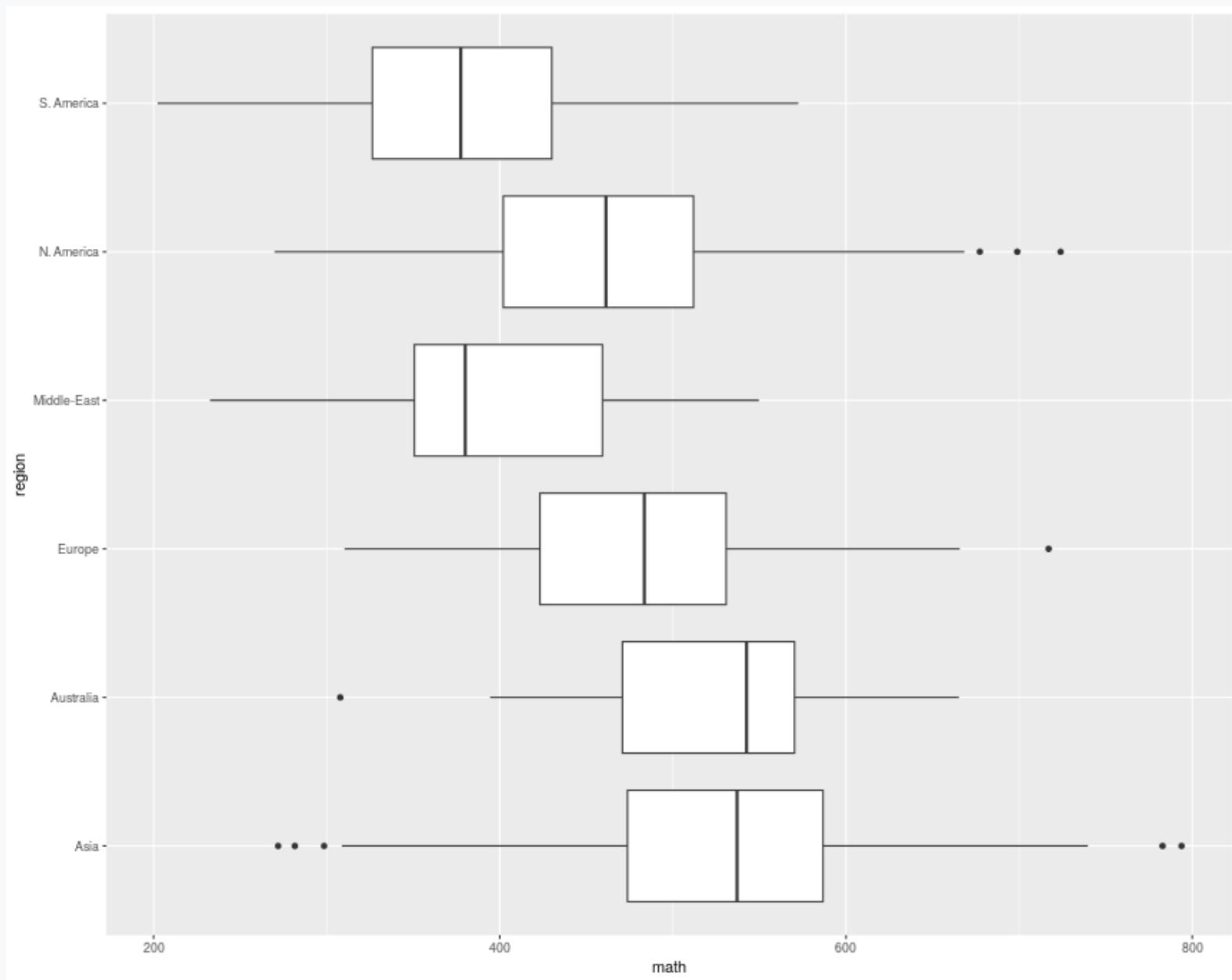
Motivating Example

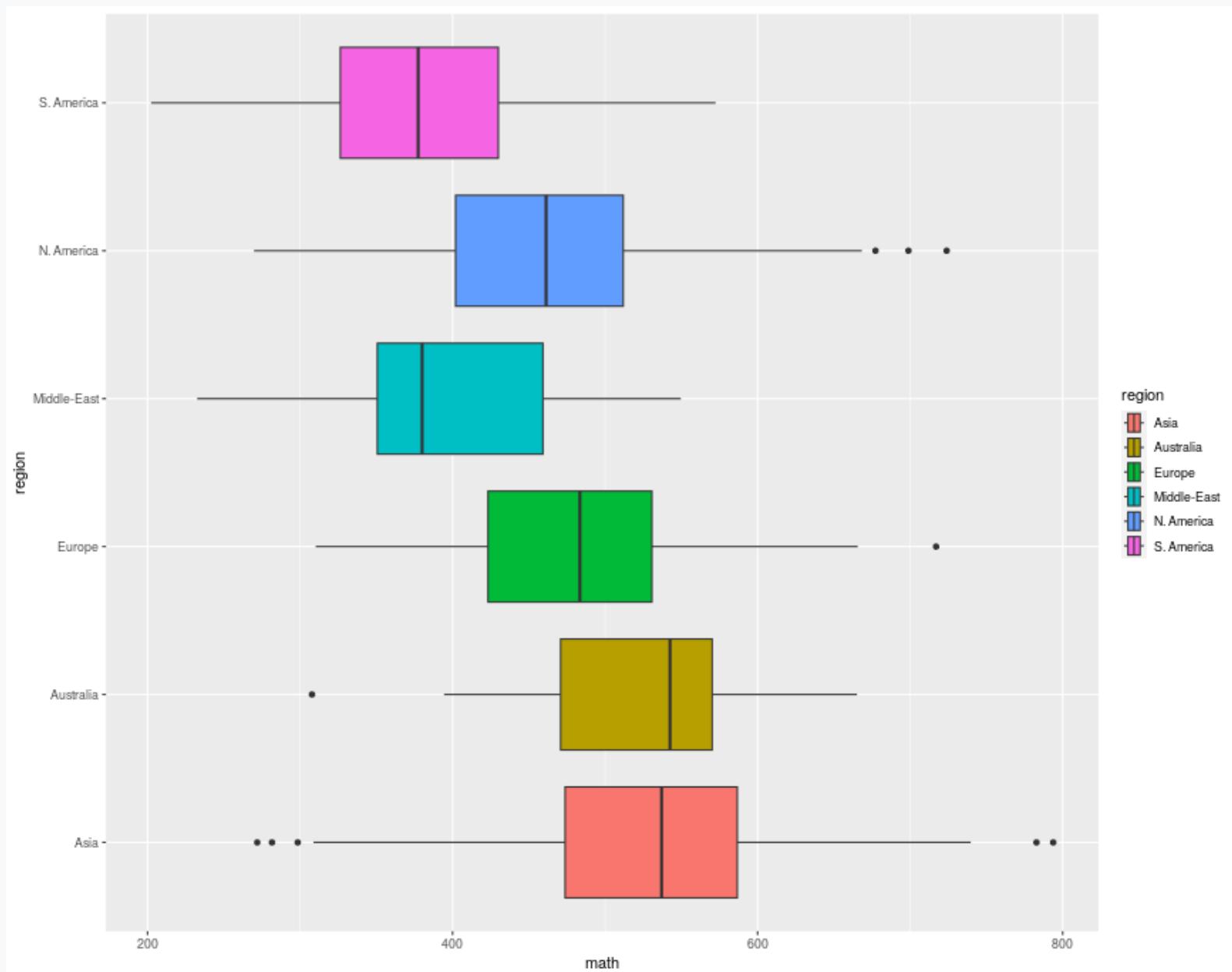


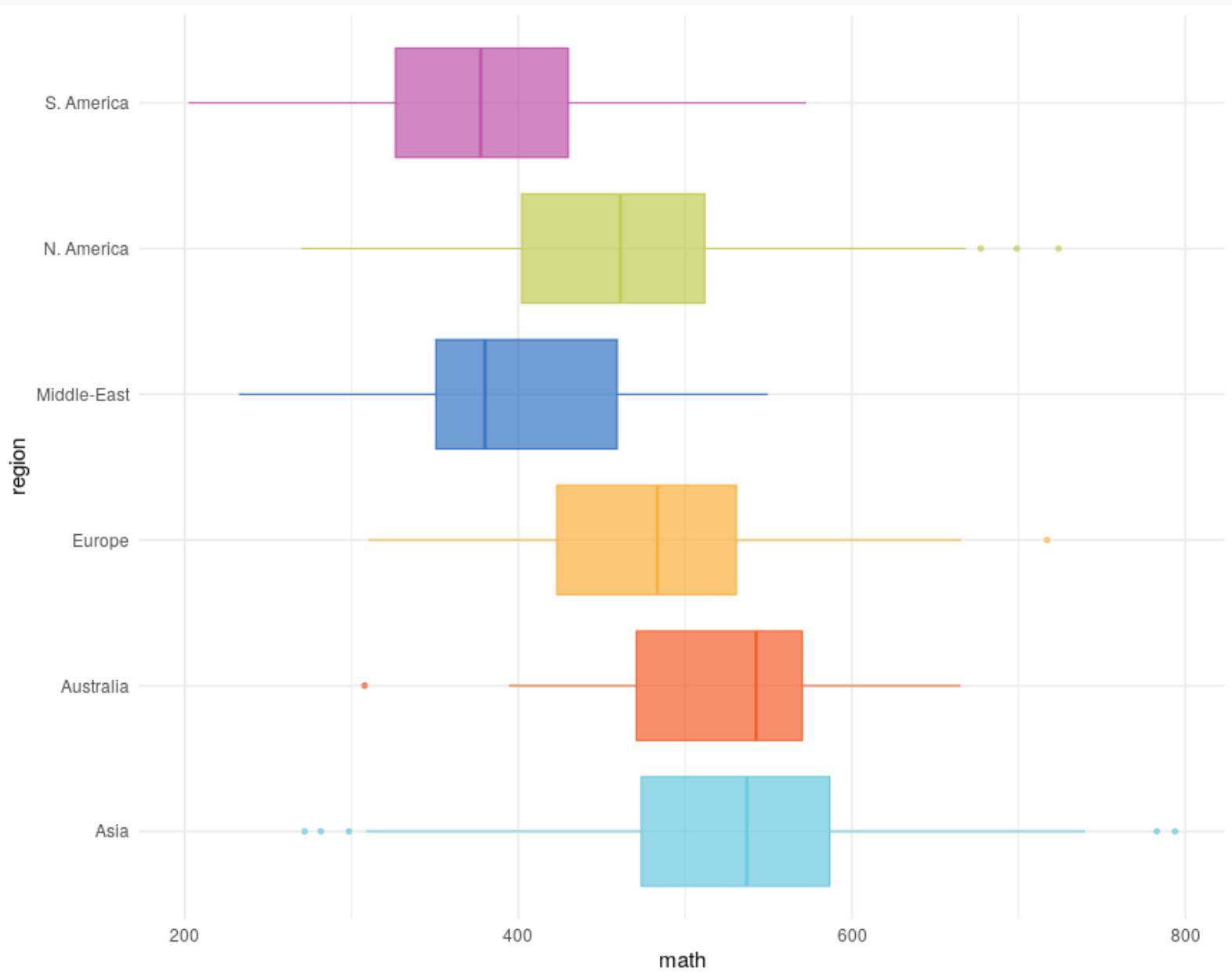
Artwork by @allison_horst

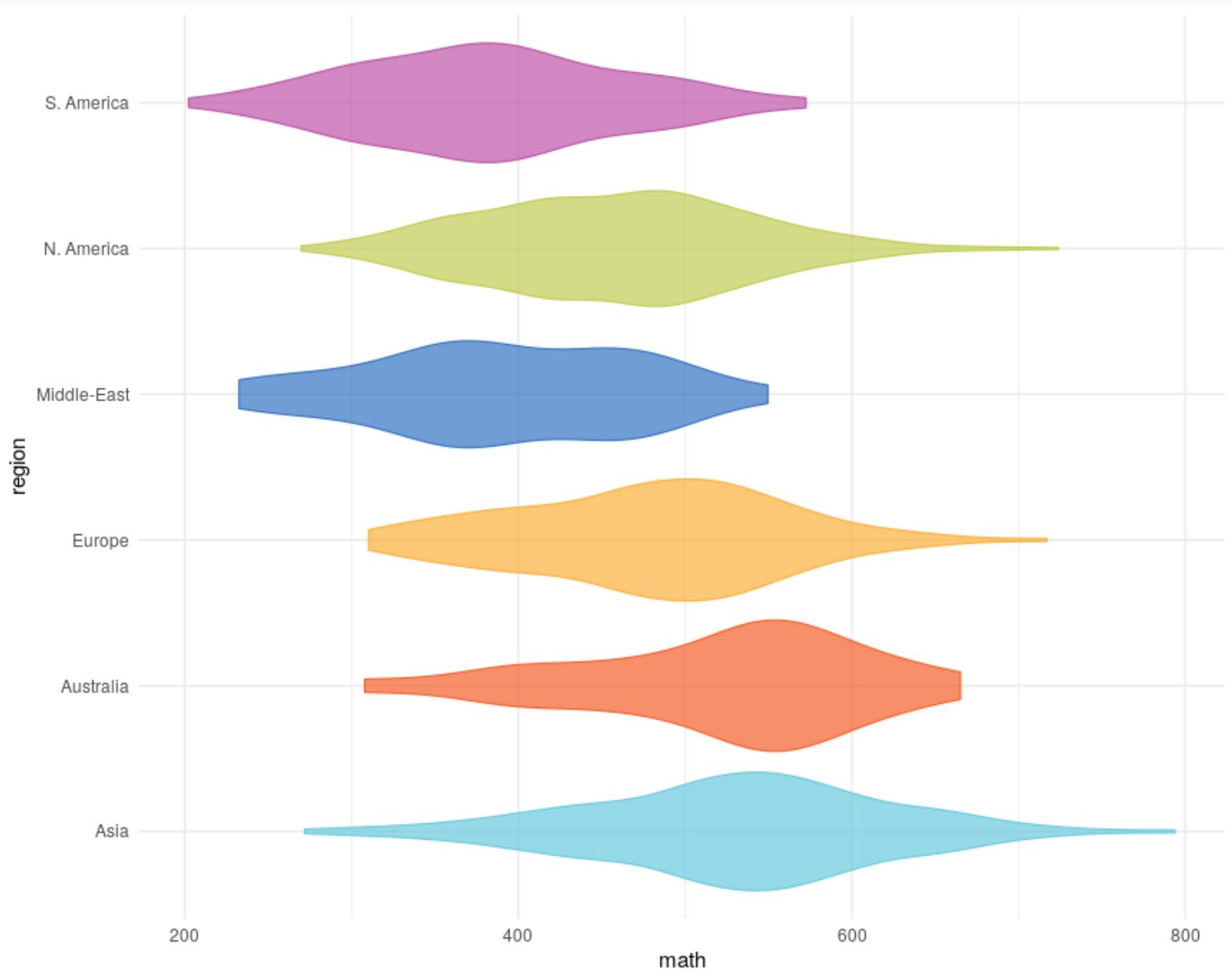


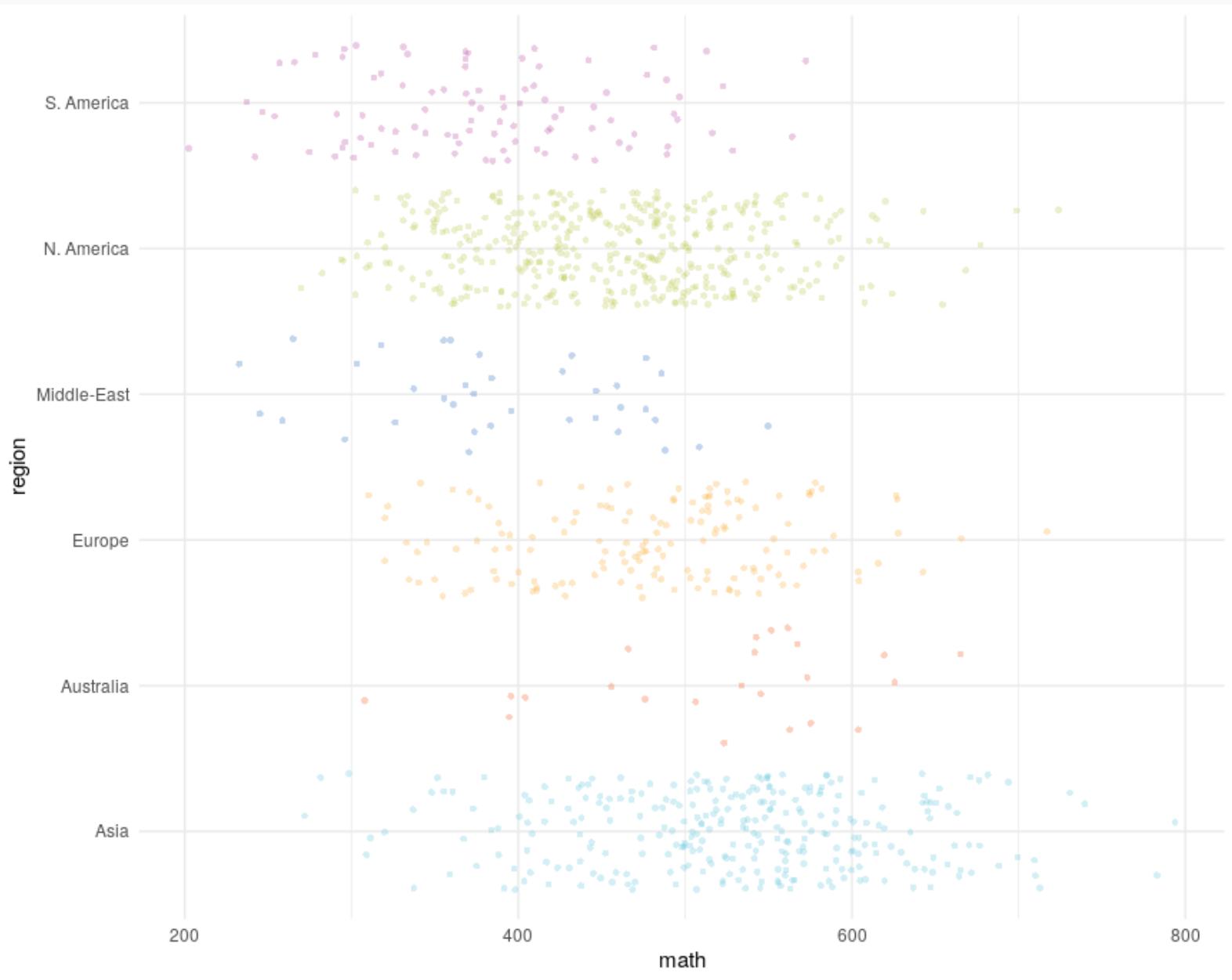


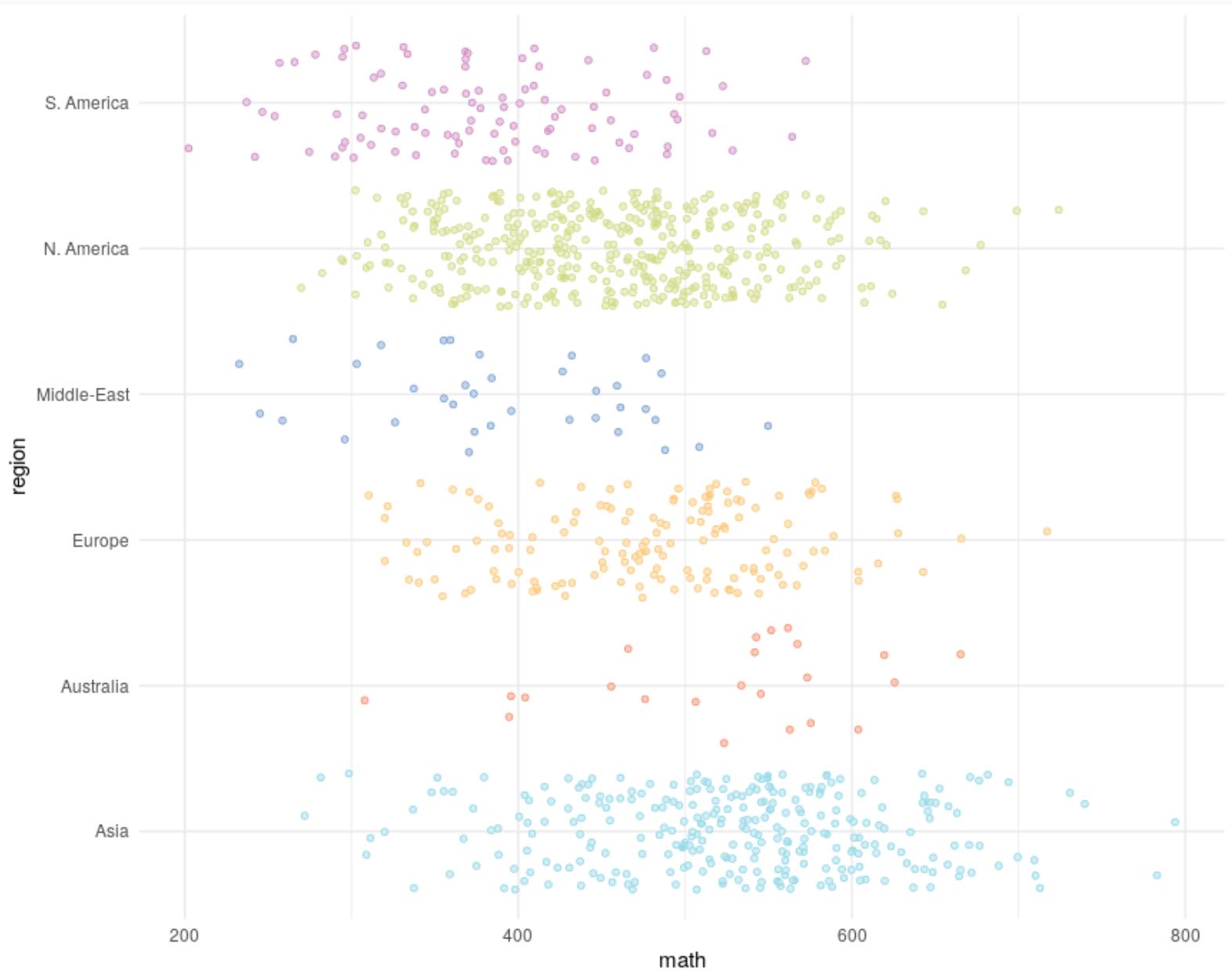








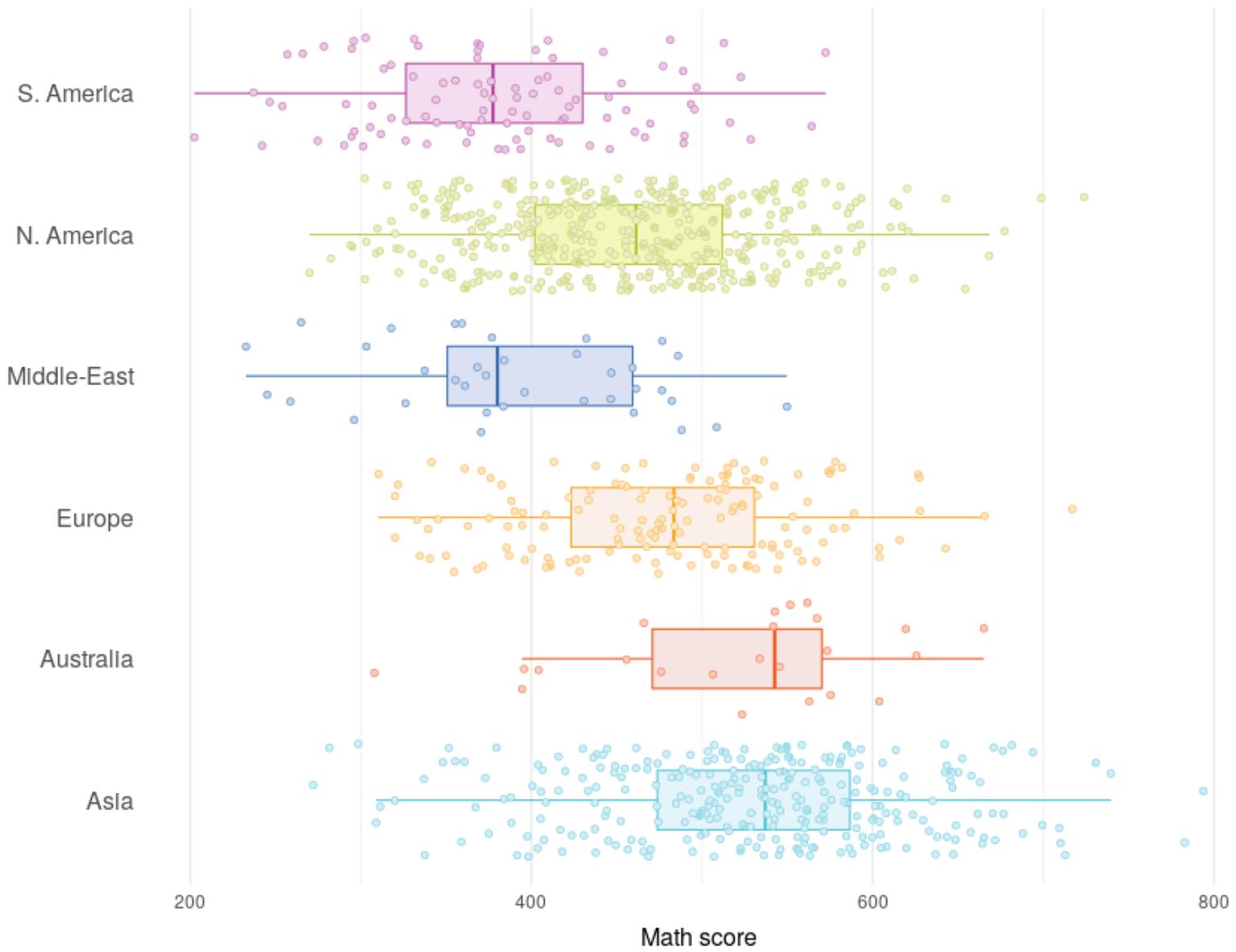




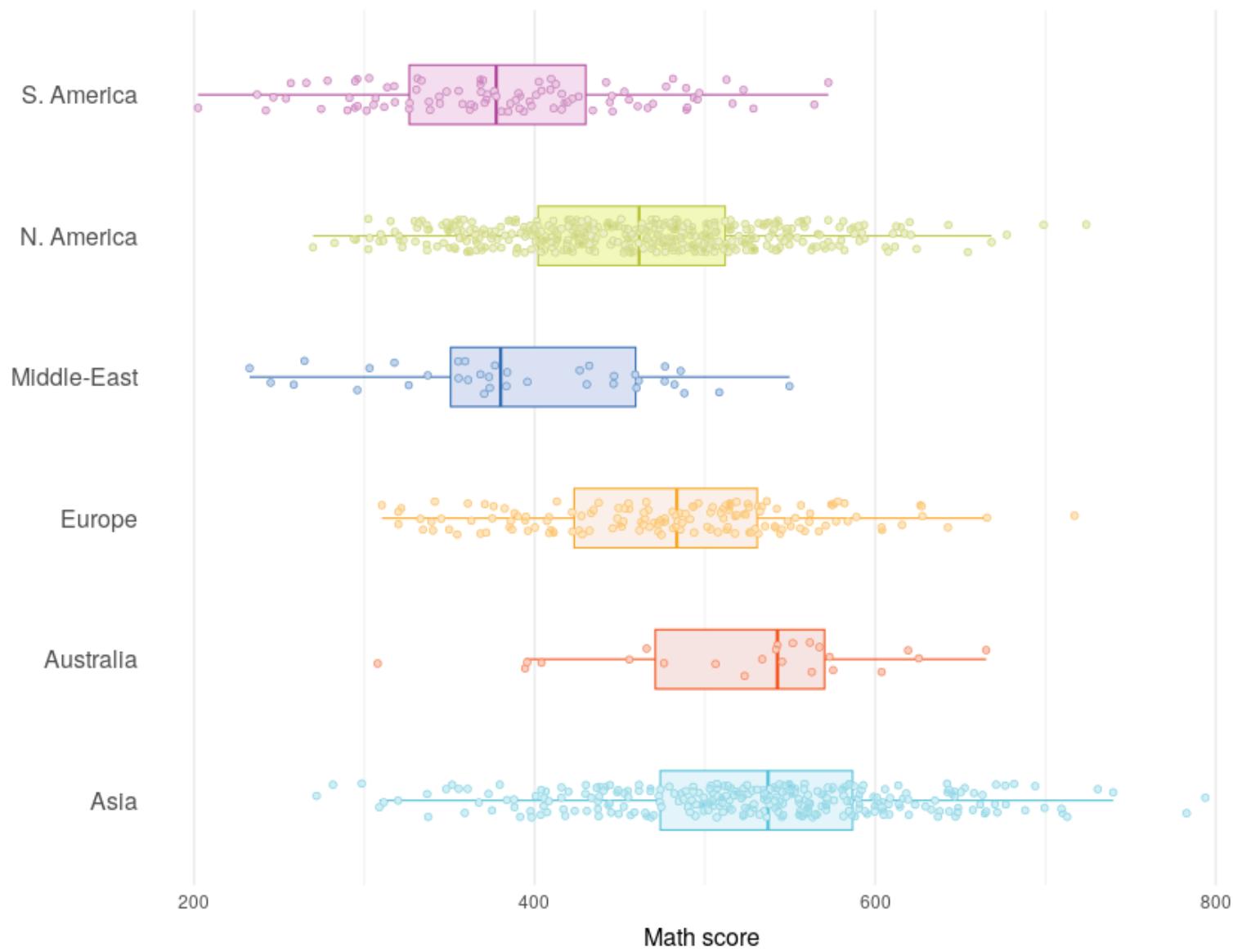
2015 PISA Math scores by region



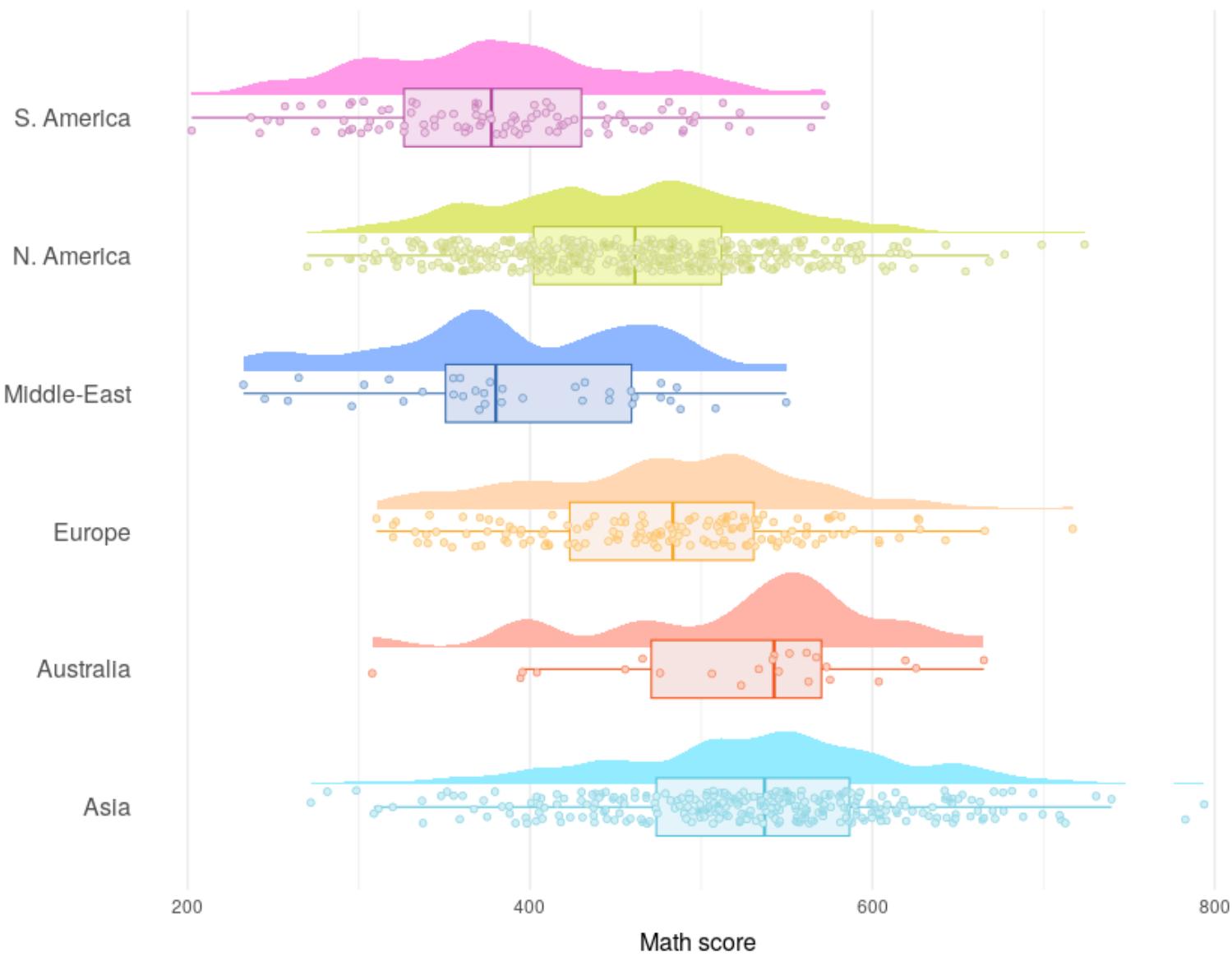
2015 PISA Math scores by region



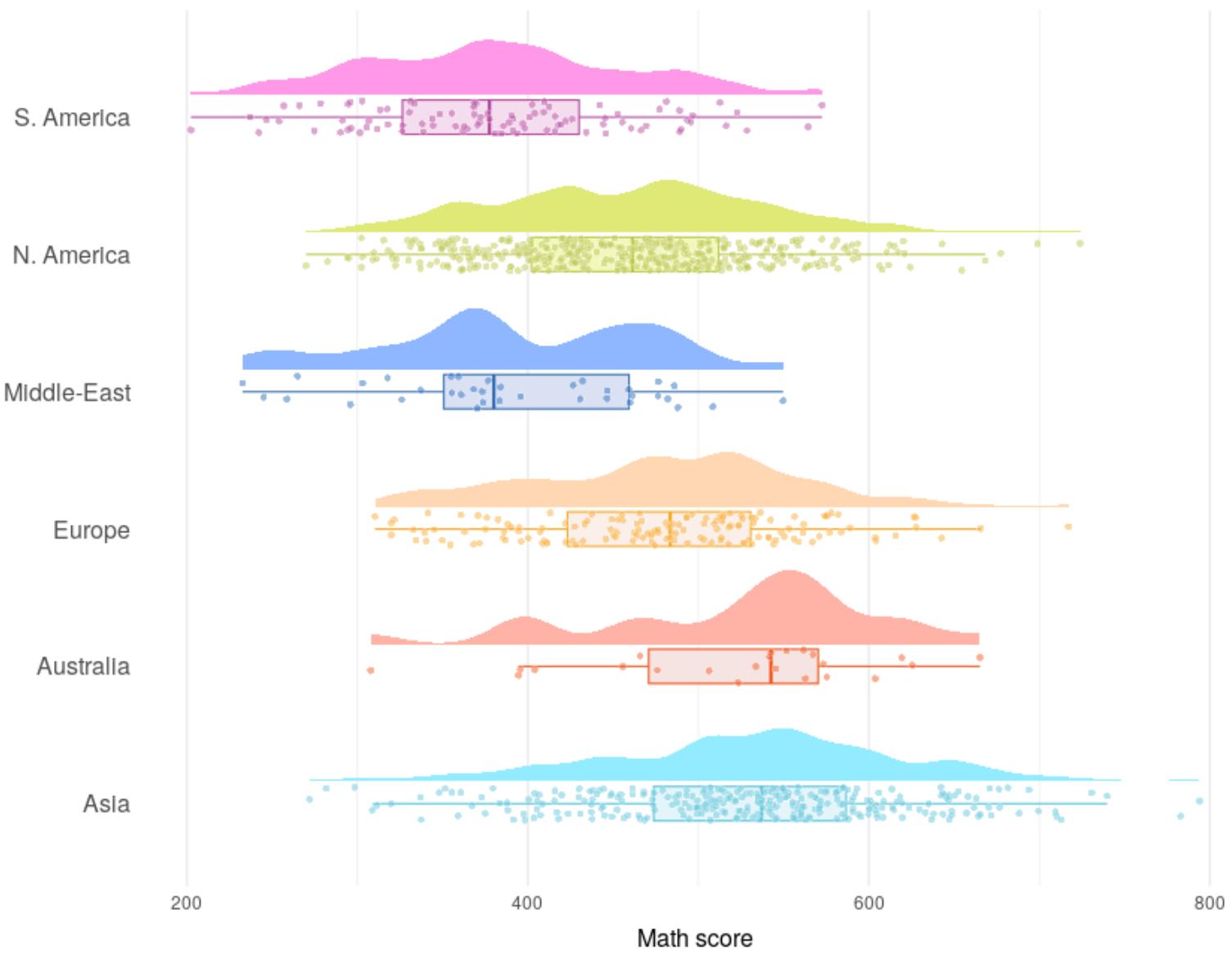
2015 PISA Math scores by region



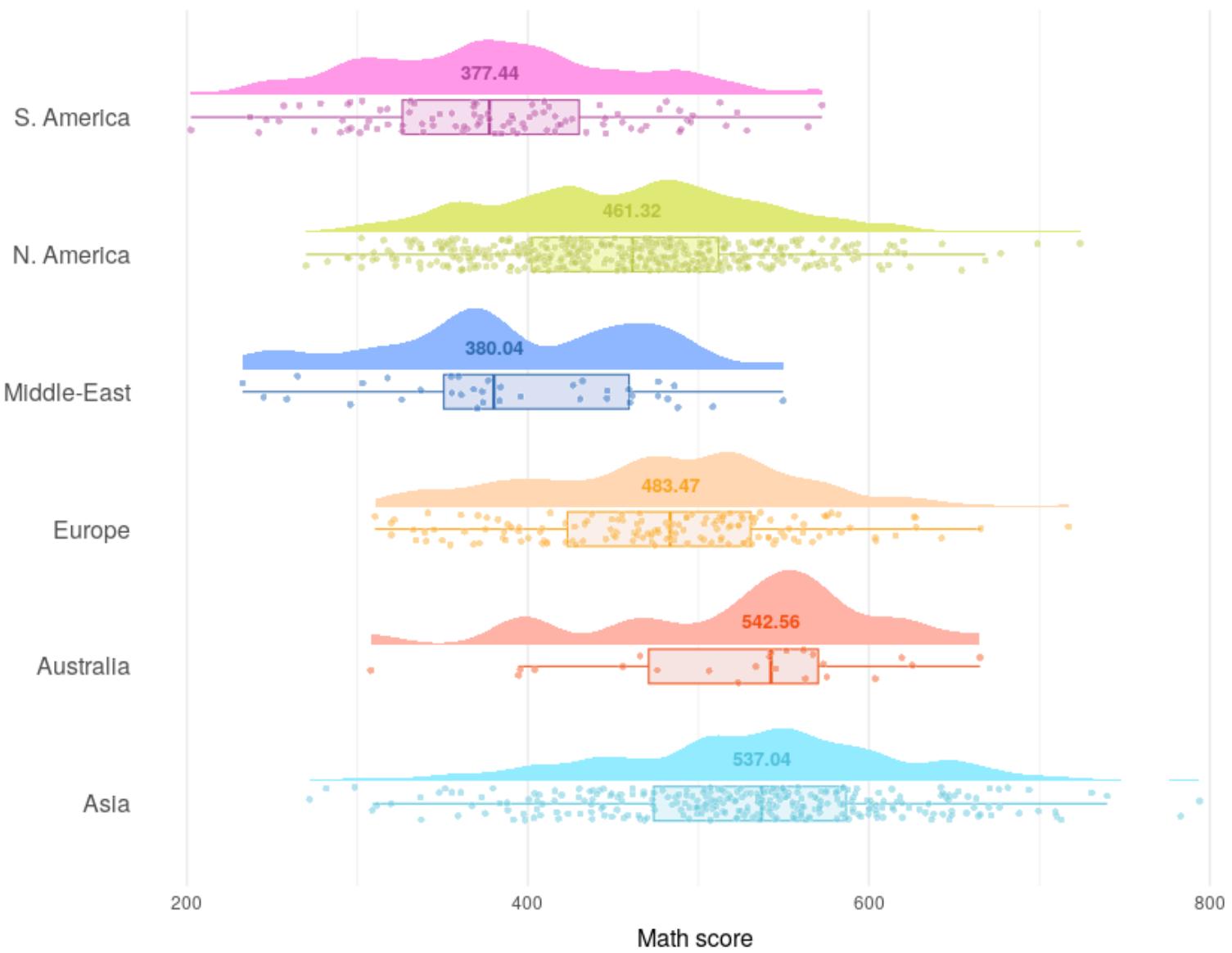
2015 PISA Math scores by region



2015 PISA Math scores by region



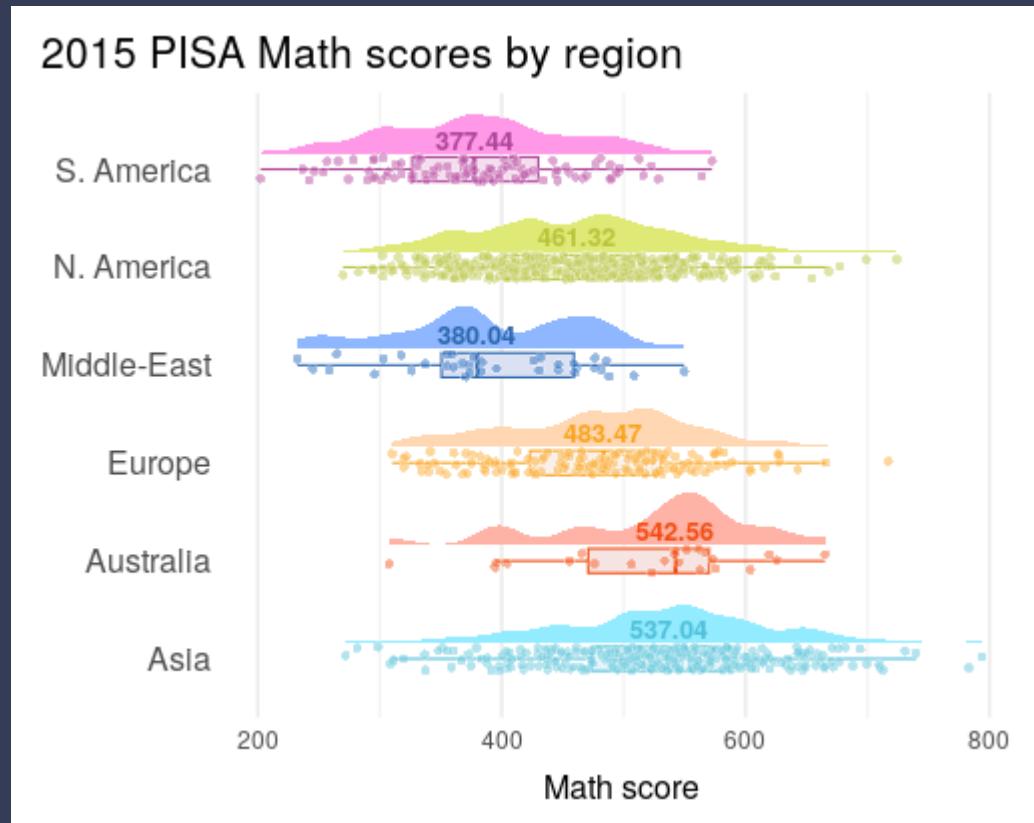
2015 PISA Math scores by region



Your Turn

Talk with your neighbor, introduce yourself, brainstorm:

- What other features would you like to include?
- What other types of visualizations would you like to see?



Resources

- Documentation: <http://ggplot2.tidyverse.org/reference/>
- RStudio cheat sheet for `ggplot2`
- Sam Tyner's [ggplot2 workshop](#)
- Thomas Lin Pedersen's ggplot2 webinar: [part 1](#) and [part 2](#)
- Cedric Scherer's "[A ggplot2 tutorial for beautiful plotting in R](#)"