

Higher dimensions

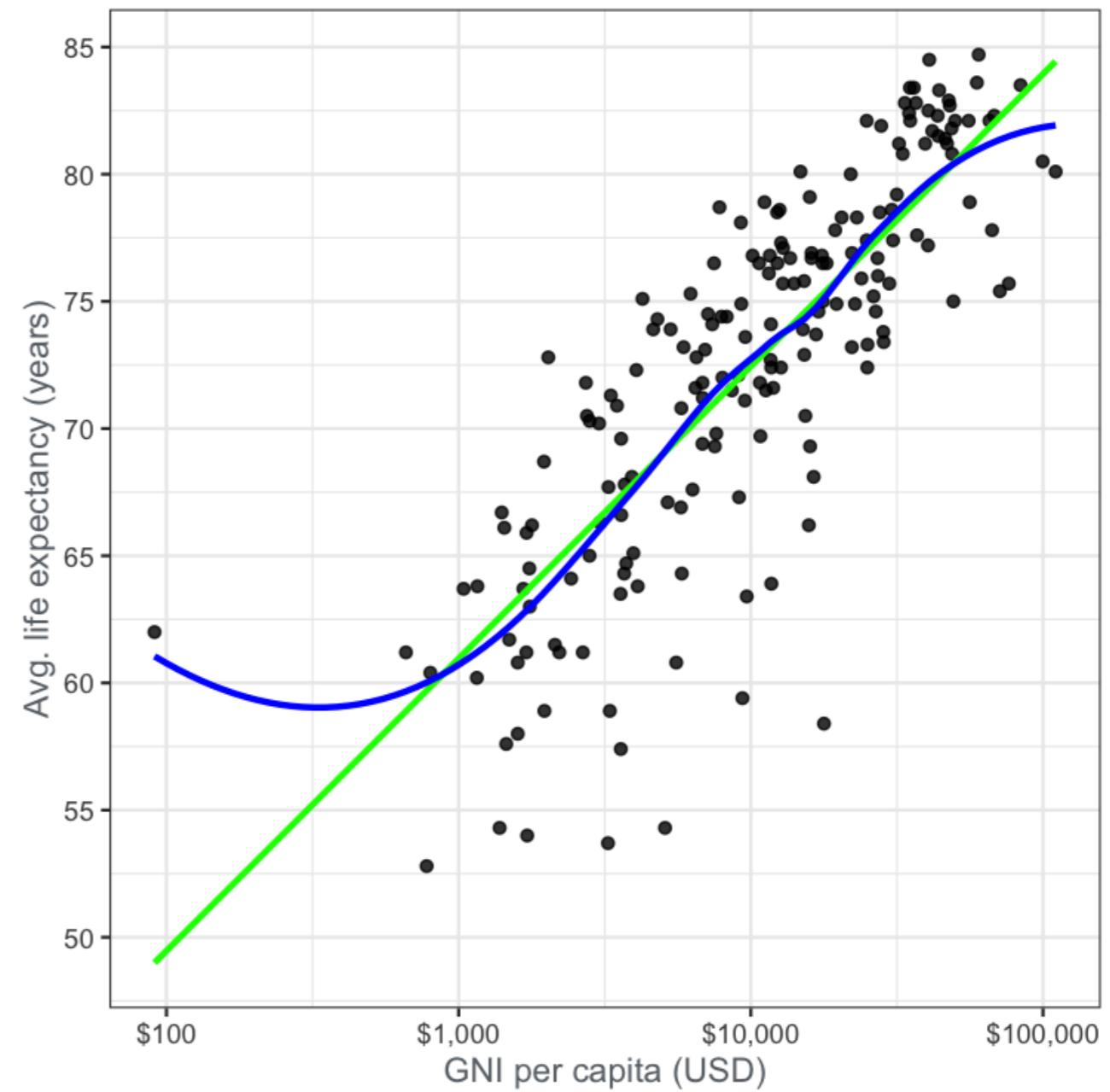
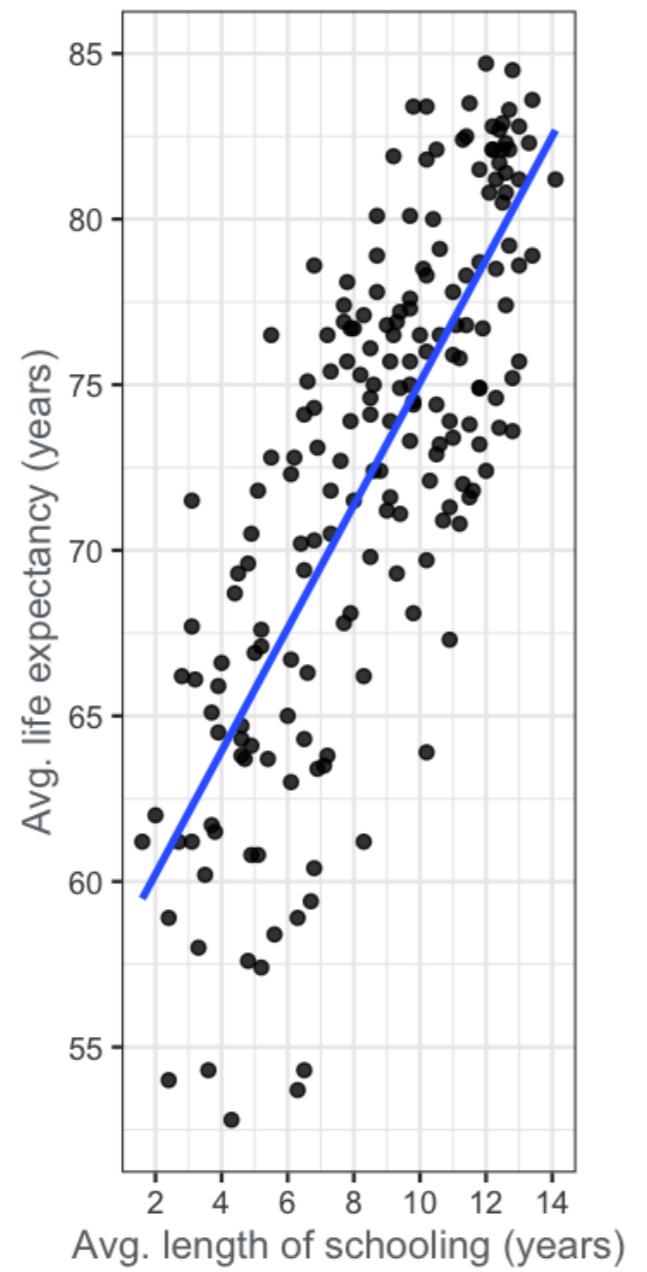
UNDERSTANDING DATA VISUALIZATION



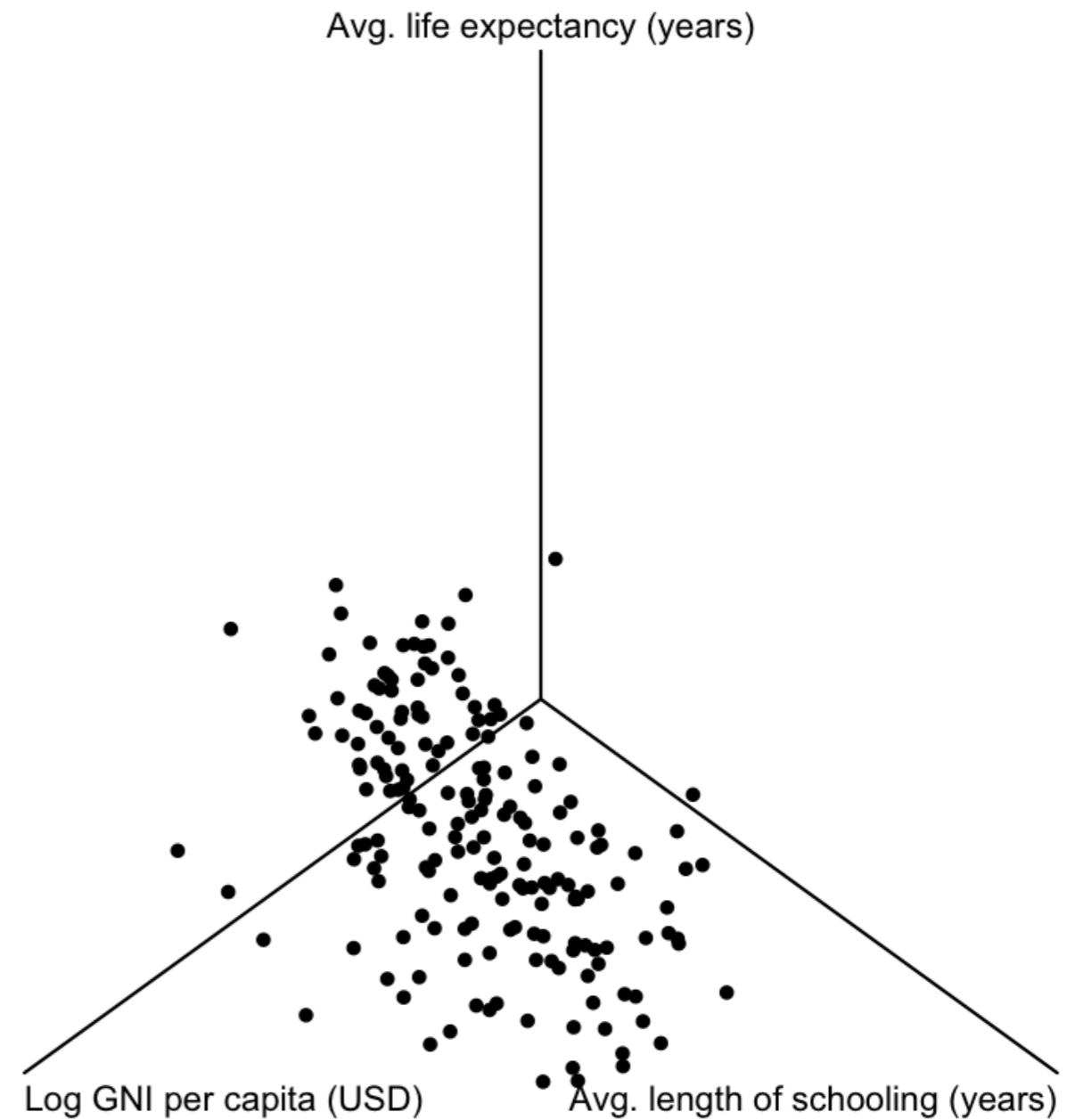
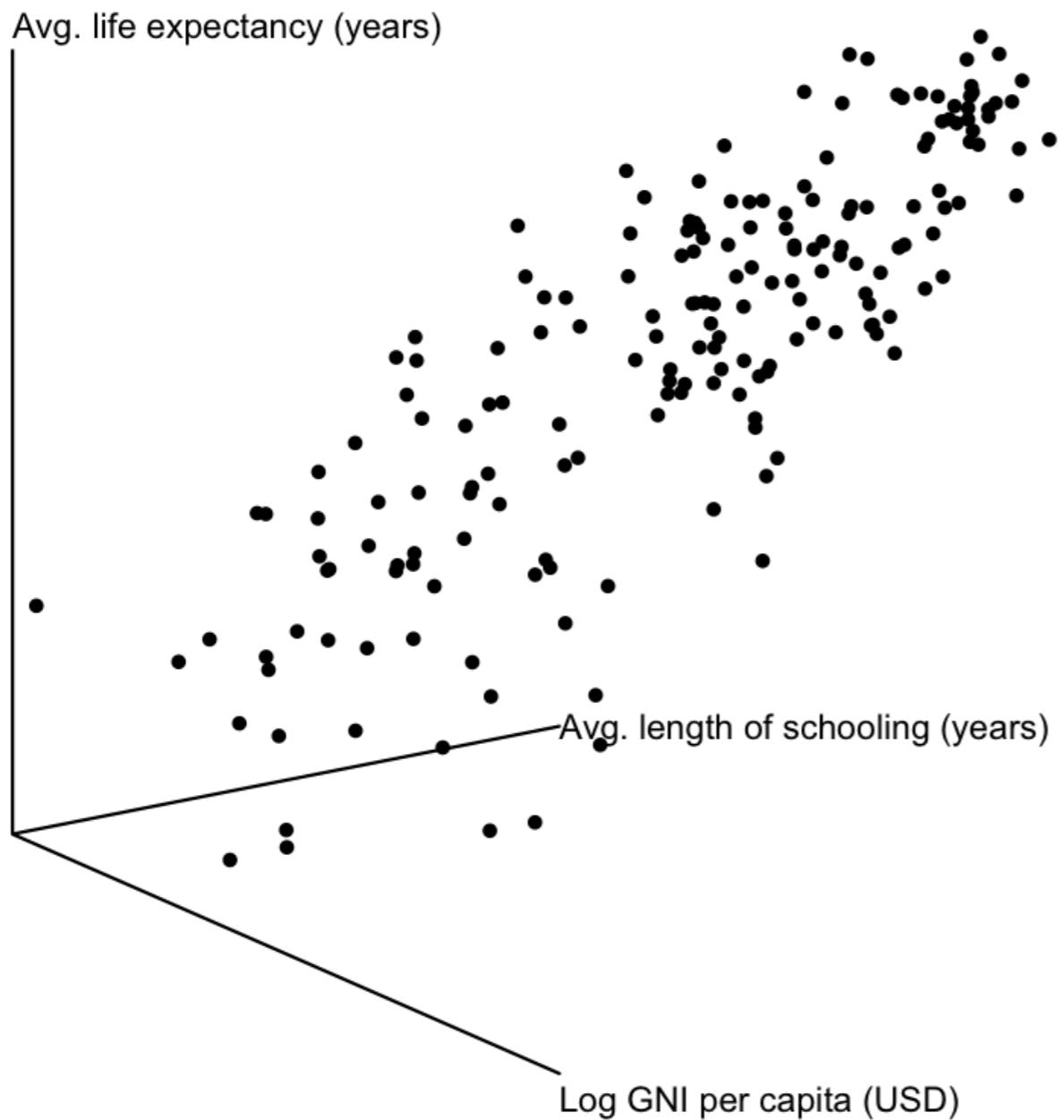
Richie Cotton

Data Evangelist at DataCamp

The UN life expectancy scatter plots



3D scatter plots

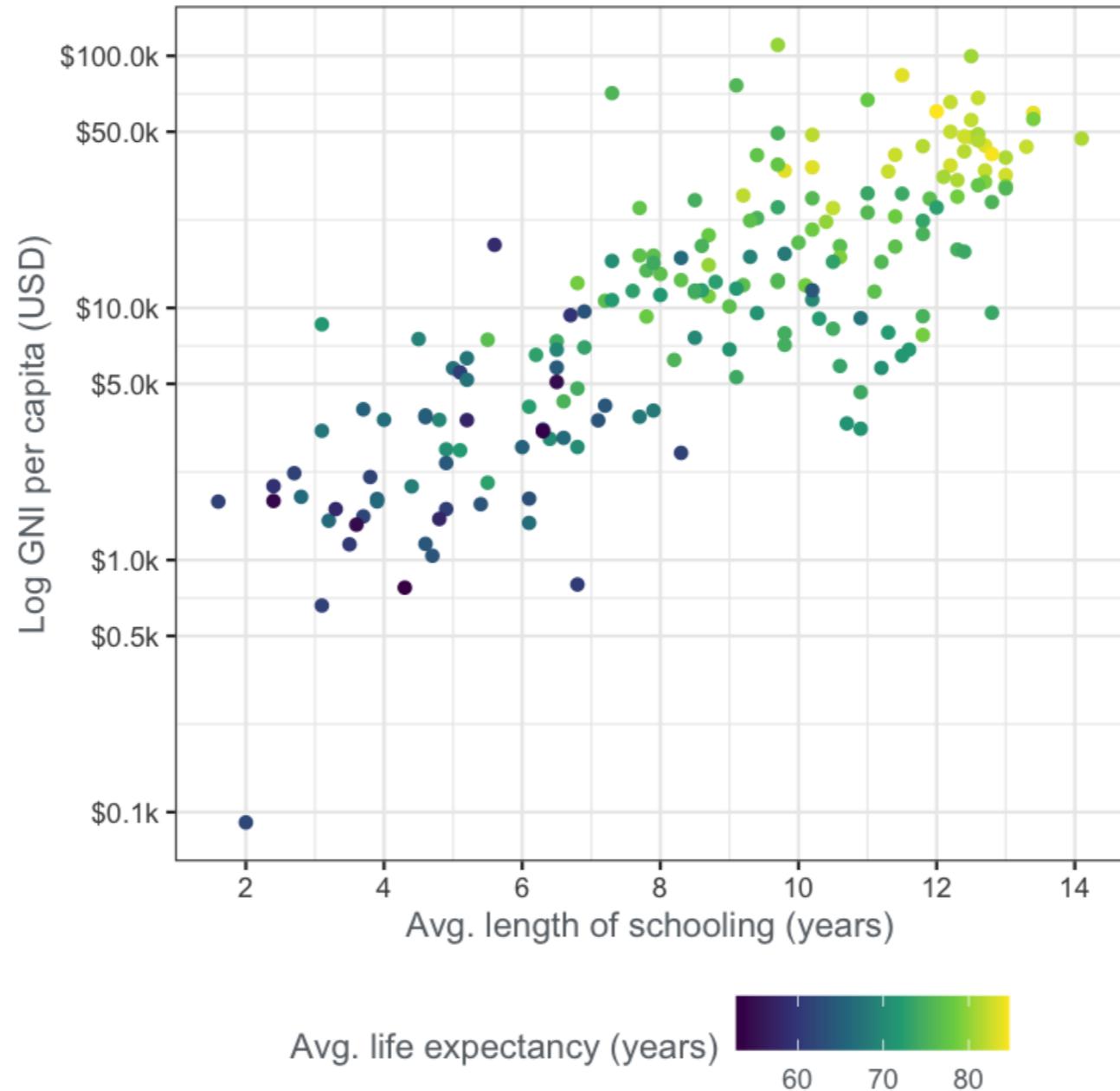


x and y are not the only dimensions

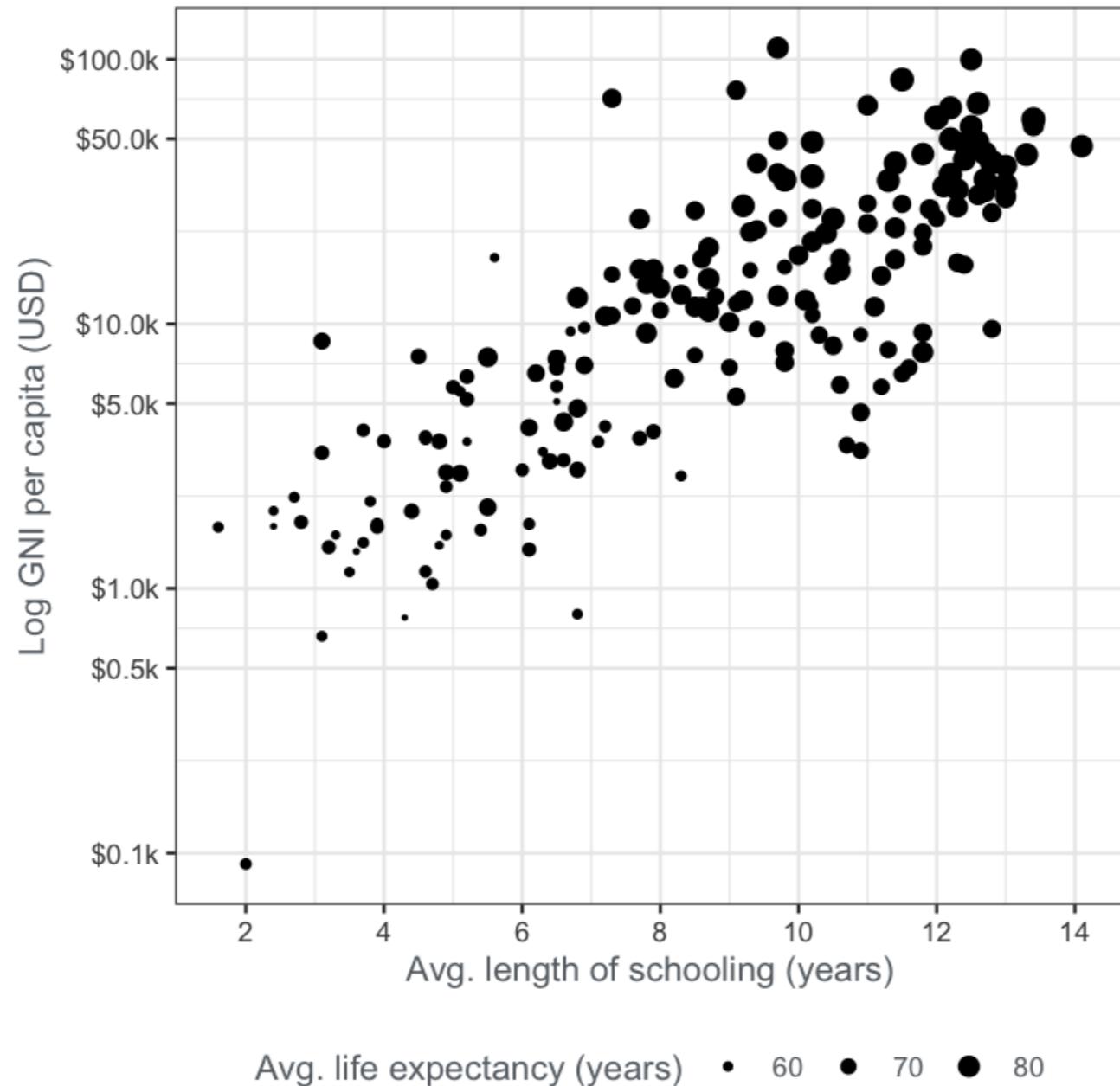
Points also have these dimensions

- color
- size
- transparency
- shape

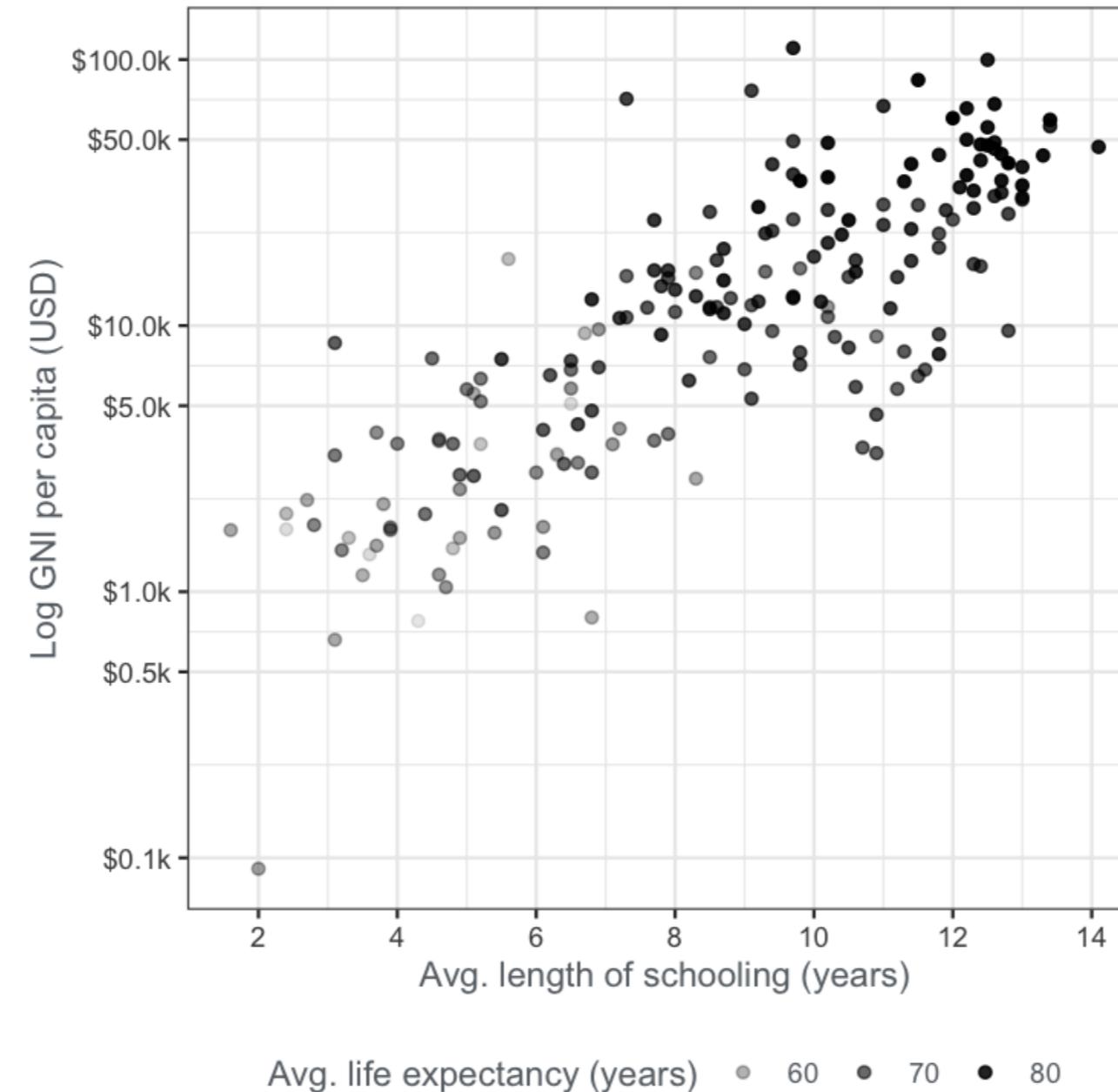
Color



Size



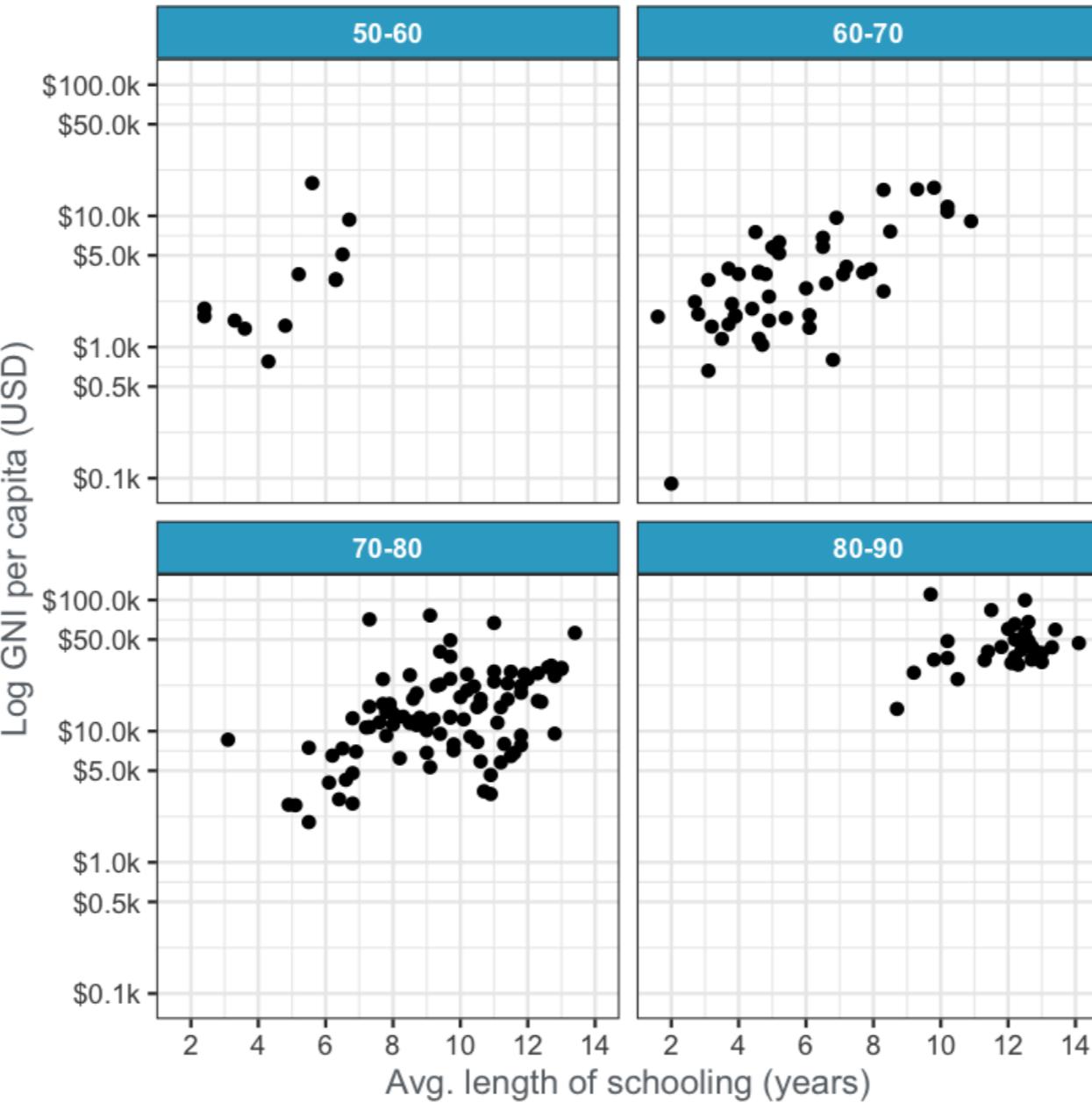
Transparency



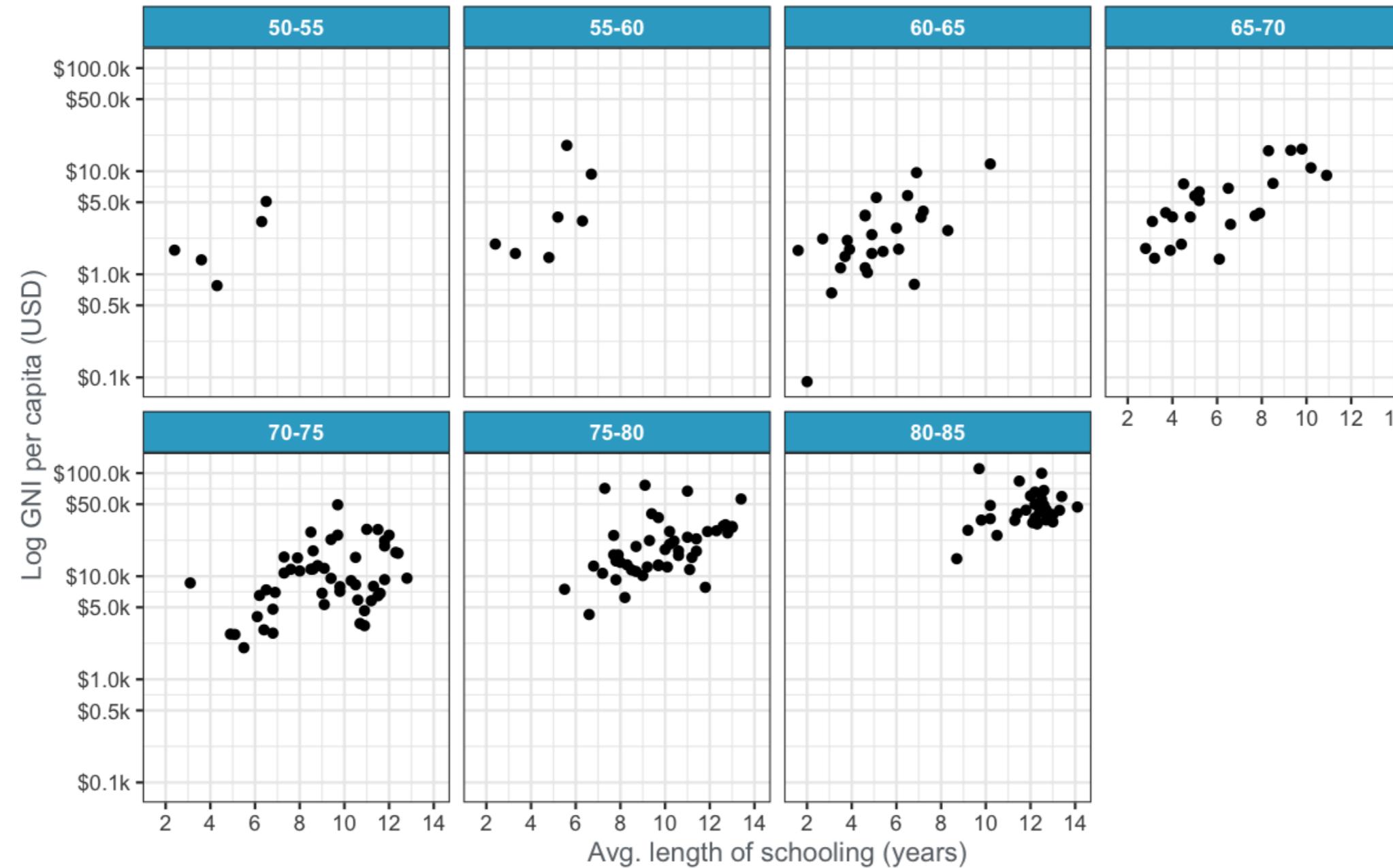
Shape



Lots of panels



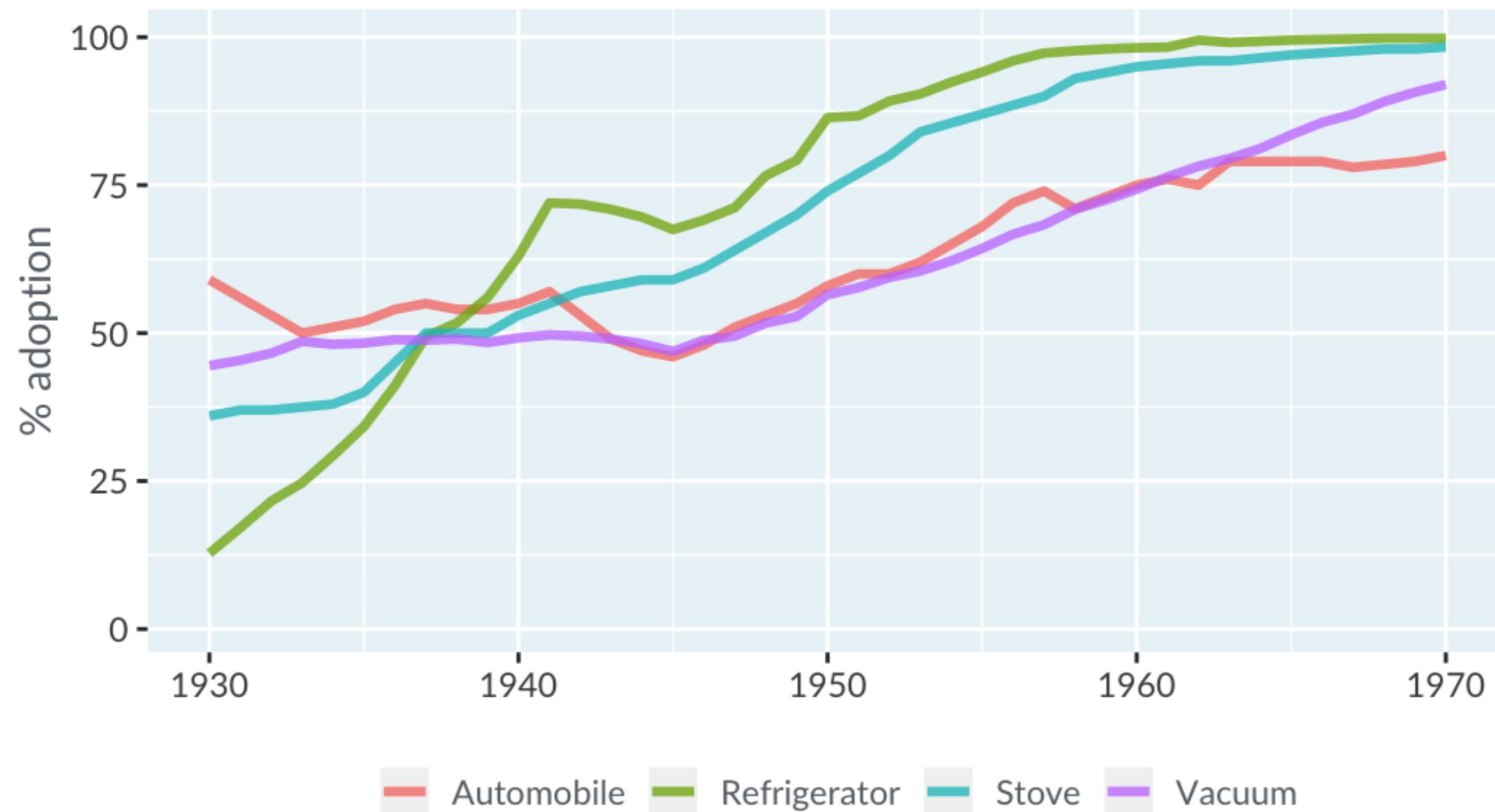
Even more panels



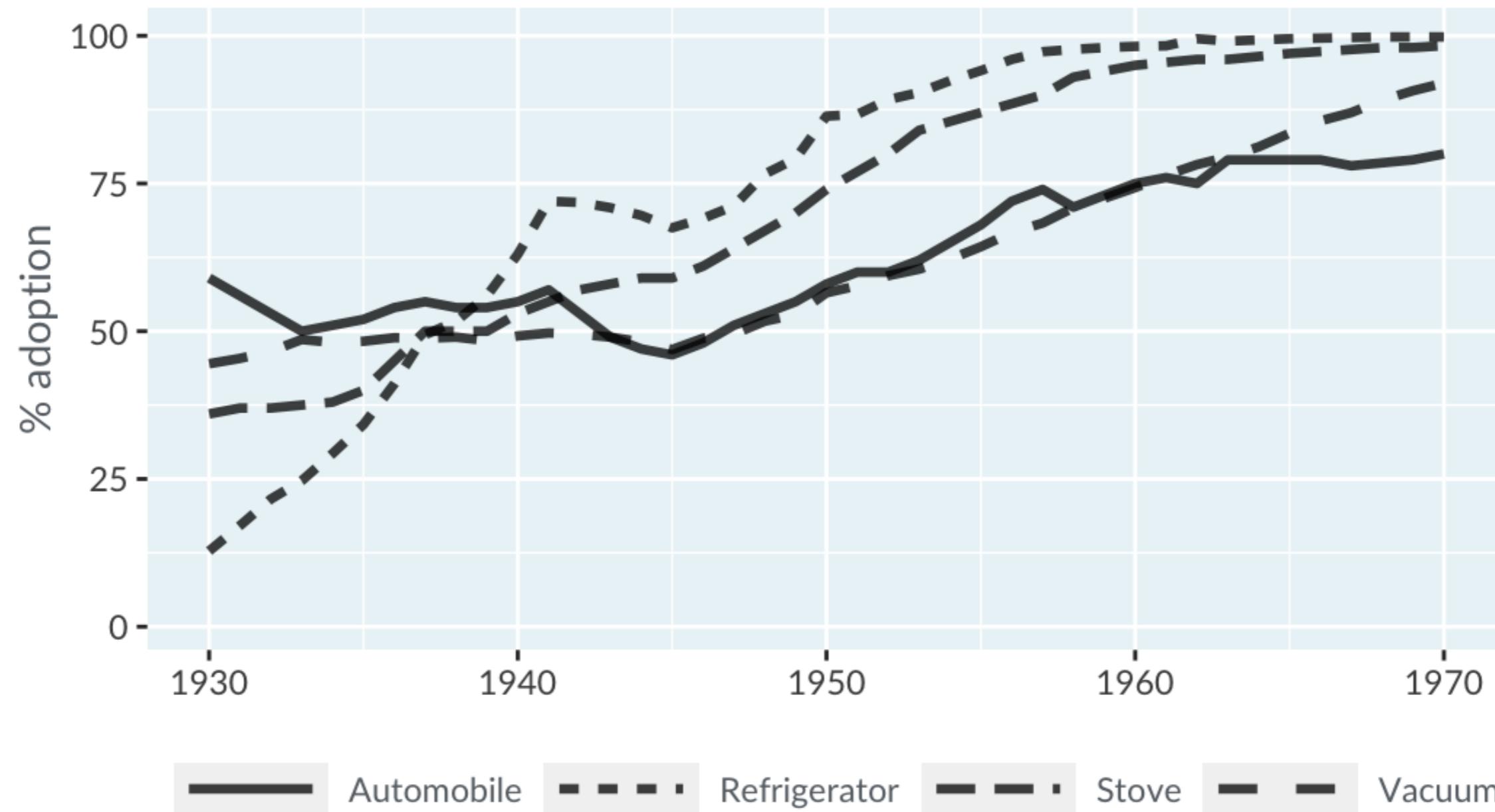
Other dimensions for line plots

- color
- thickness
- transparency
- line type (solid, dashes, dots)

Color



Linetype



Let's practice!

UNDERSTANDING DATA VISUALIZATION

Using color

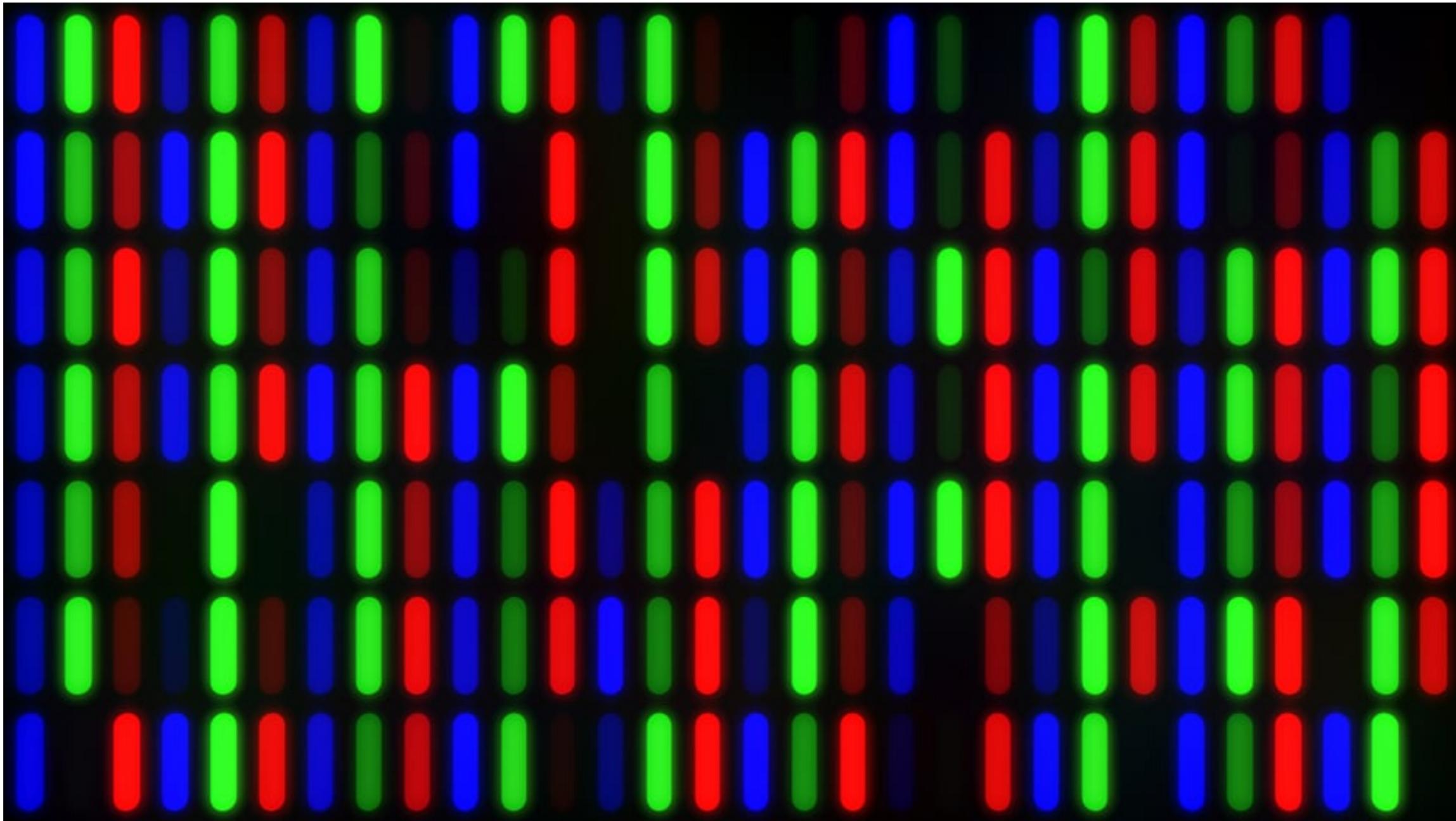
UNDERSTANDING DATA VISUALIZATION



Richie Cotton

Data Evangelist at DataCamp

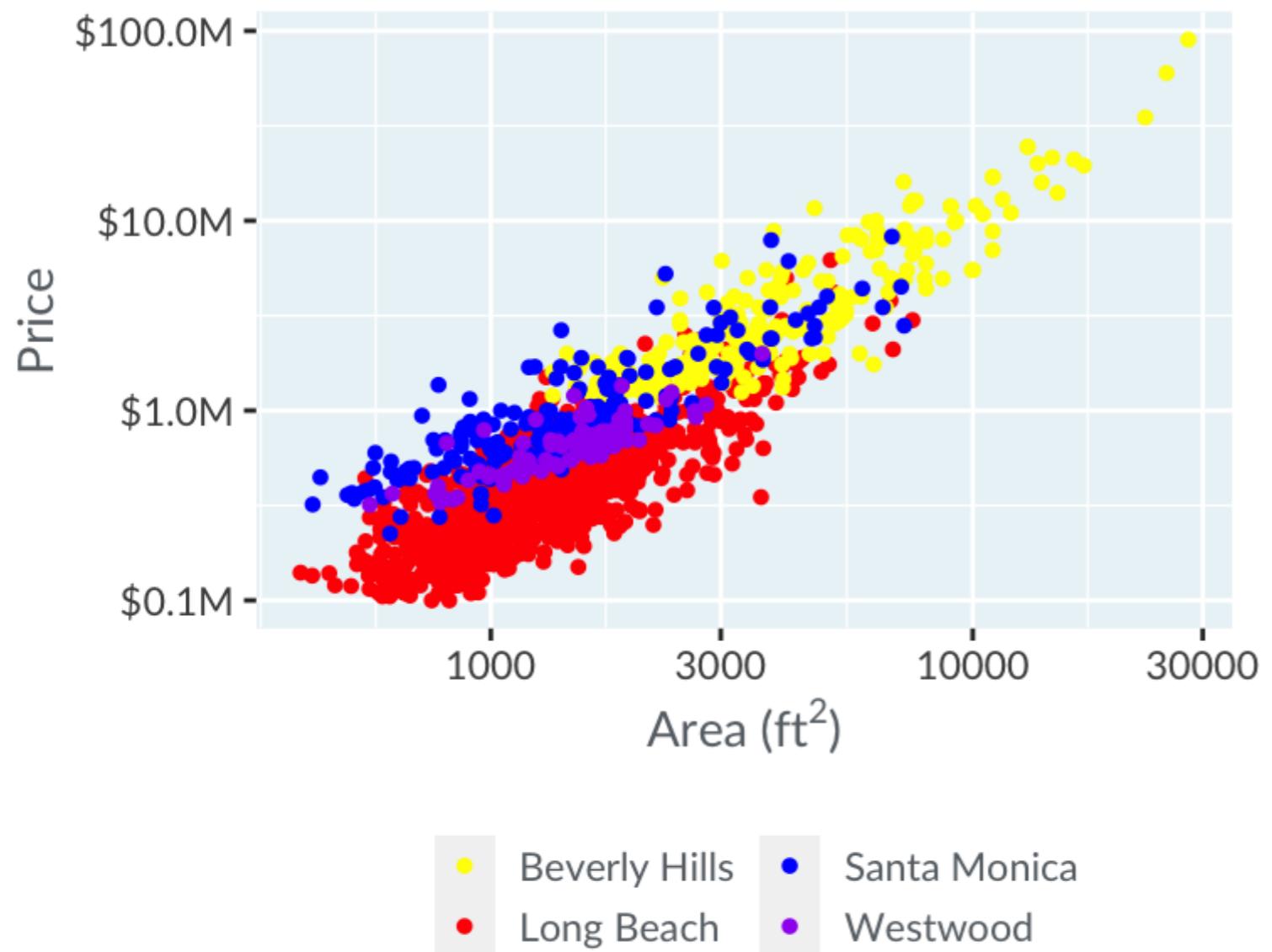
Colorspaces: Red-Green-Blue



Colorspaces: Cyan-Magenta-Yellow-blackK



Choosing a plotting palette



- Usually, each color should stand out as much as other colors.
- The perceptual distance from one color in the plot to the next should be constant.

Colorspaces: Hue-Chroma-Luminance

Hue



Colorspaces: Hue-Chroma-Luminance

Hue



Chroma (green)



Chroma (magenta)



Colorspaces: Hue-Chroma-Luminance

Hue



Chroma (green)



Chroma (magenta)



Luminance (cyan)



Luminance (red)

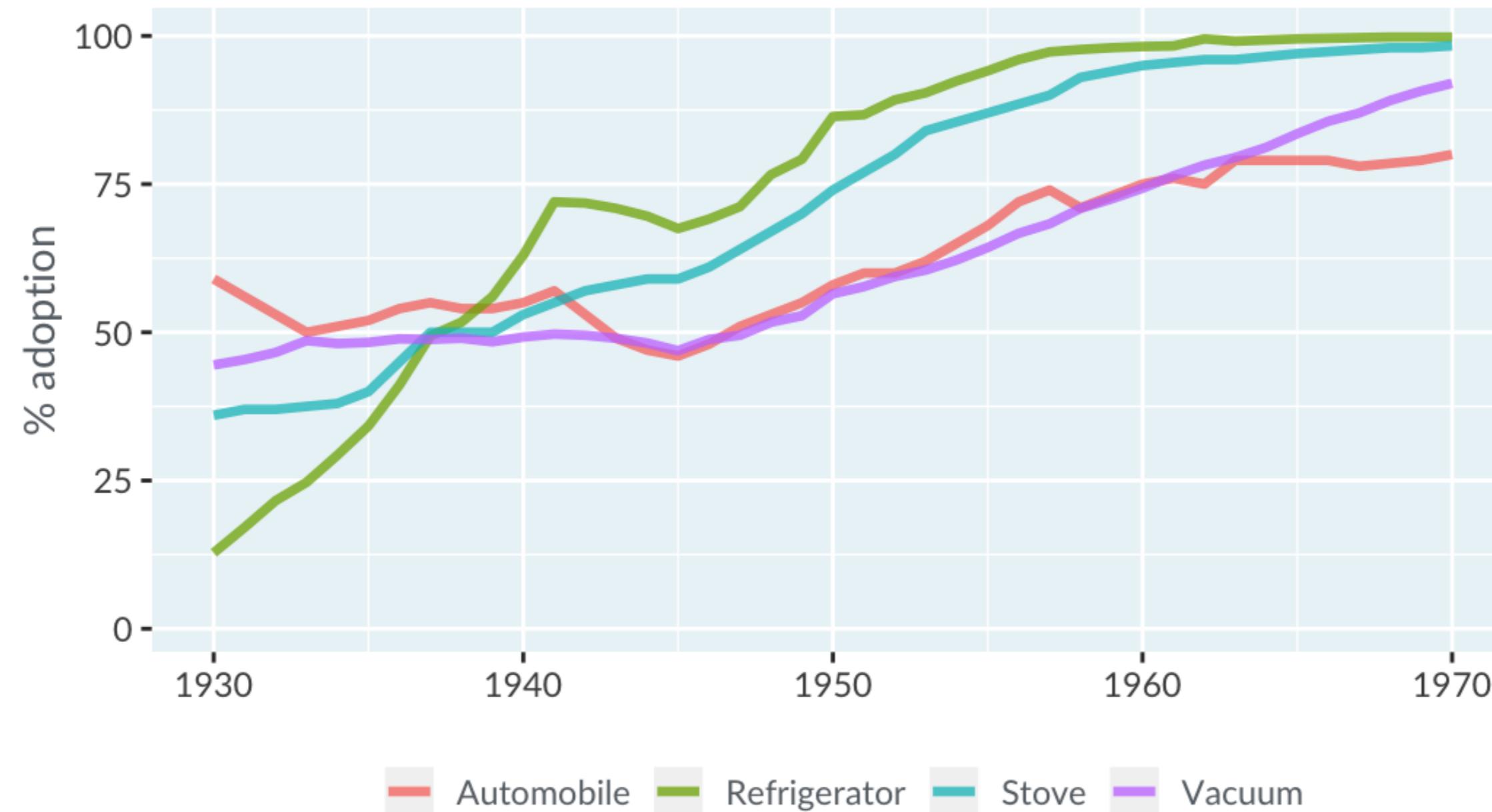


Three types of color scale: qualitative

Type	Purpose	What to vary
qualitative	Distinguish unordered categories	hue



Qualitative palette example

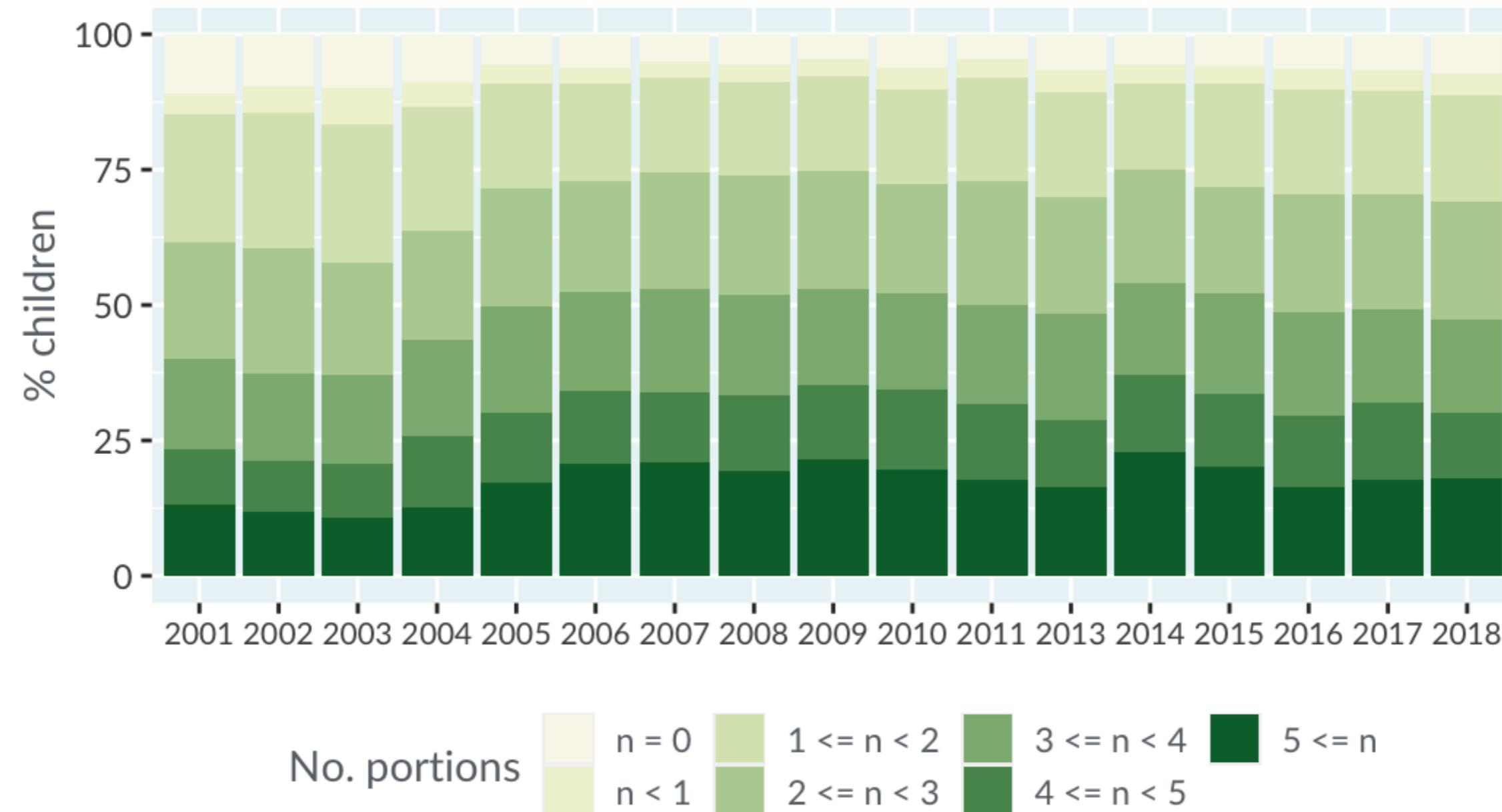


Three types of color scale: sequential

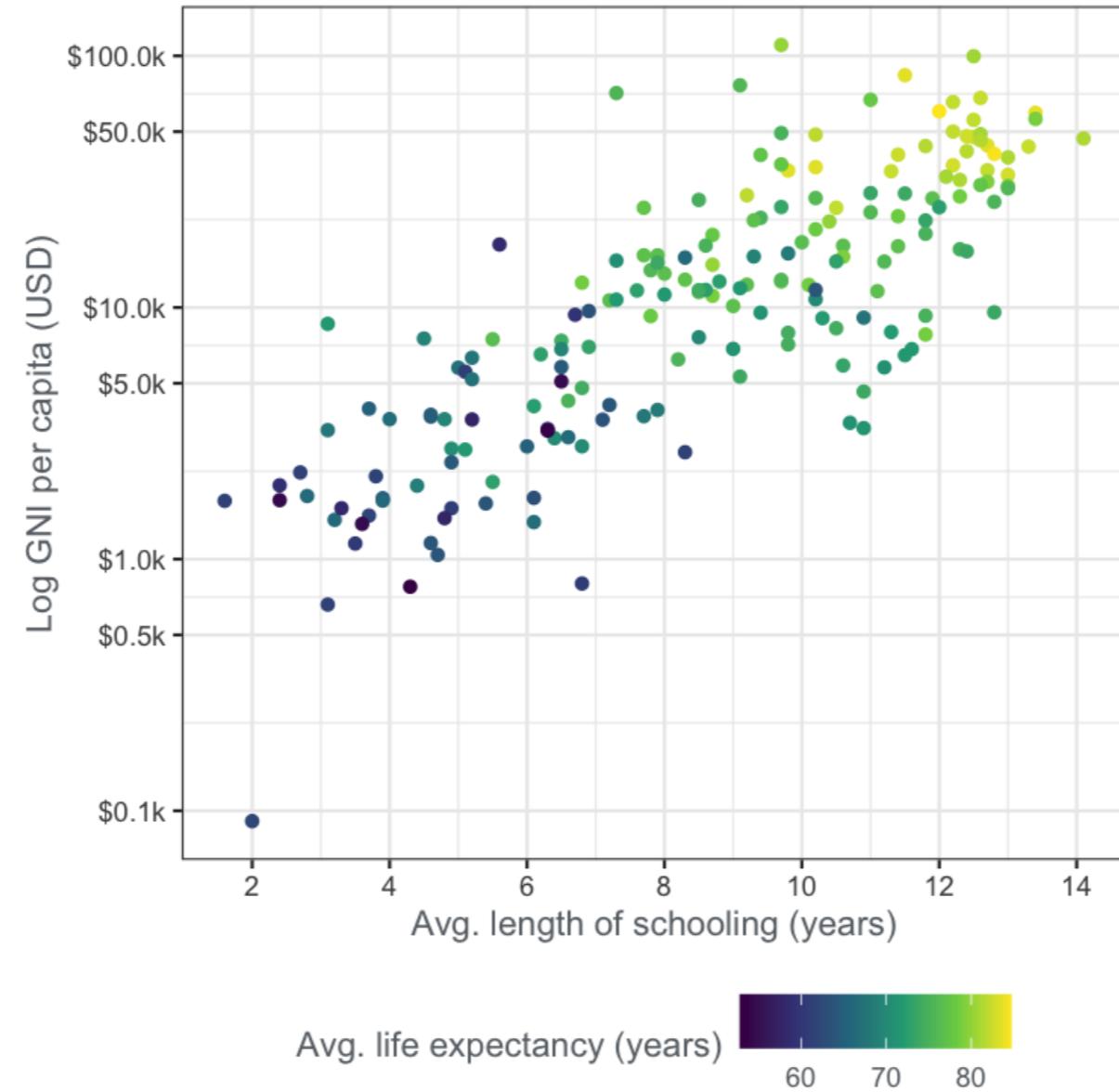
Type	Purpose	What to vary
sequential	Show ordering	chroma or luminance



Sequential palette example



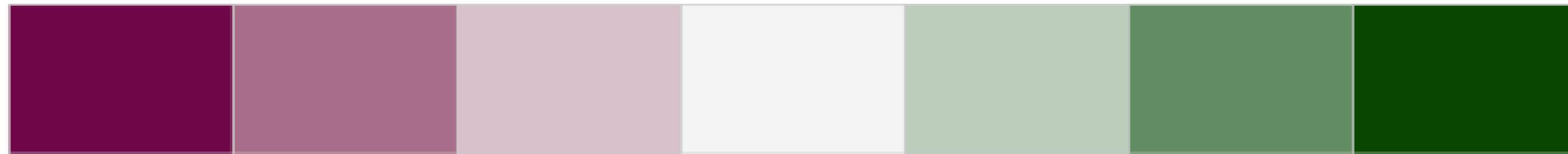
Another sequential palette example



¹ Viridis color scale: <https://bids.github.io/colormap>

Three types of color scale: diverging

Type	Purpose	What to vary
diverging	Show above or below a midpoint	chroma or luminance, with 2 hues

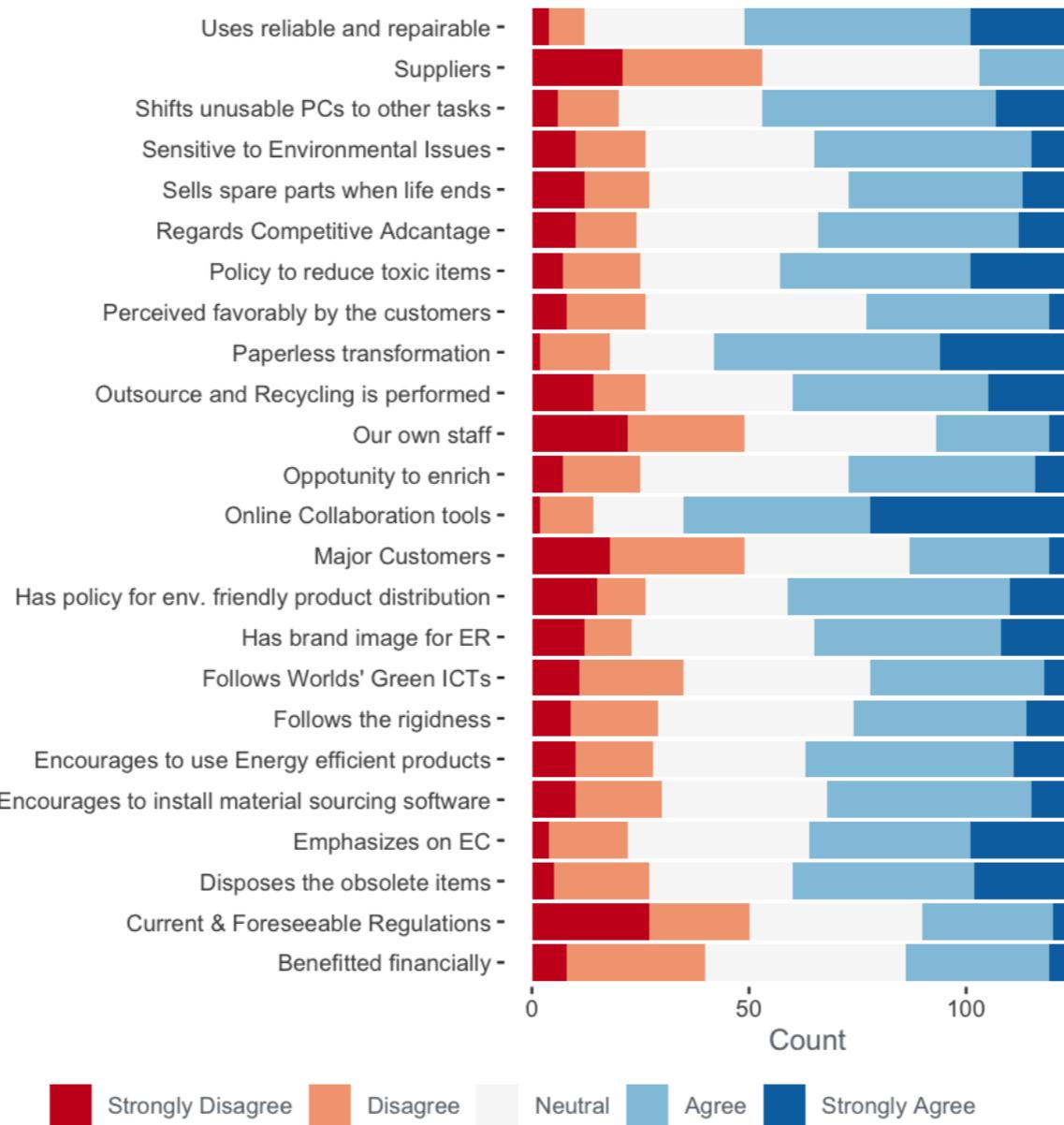


Green Tech in Malaysia survey dataset

question	response	n
Uses reliable and repairable	Strongly Disagree	4
Uses reliable and repairable	Disagree	8
Uses reliable and repairable	Neutral	37
Uses reliable and repairable	Agree	52
Uses reliable and repairable	Strongly Agree	26
...

¹ Islam et al. (2019) <http://dx.doi.org/10.17632/wggvryfhsk.1>

Diverging palette example



Let's practice!

UNDERSTANDING DATA VISUALIZATION

Plotting many variables at once

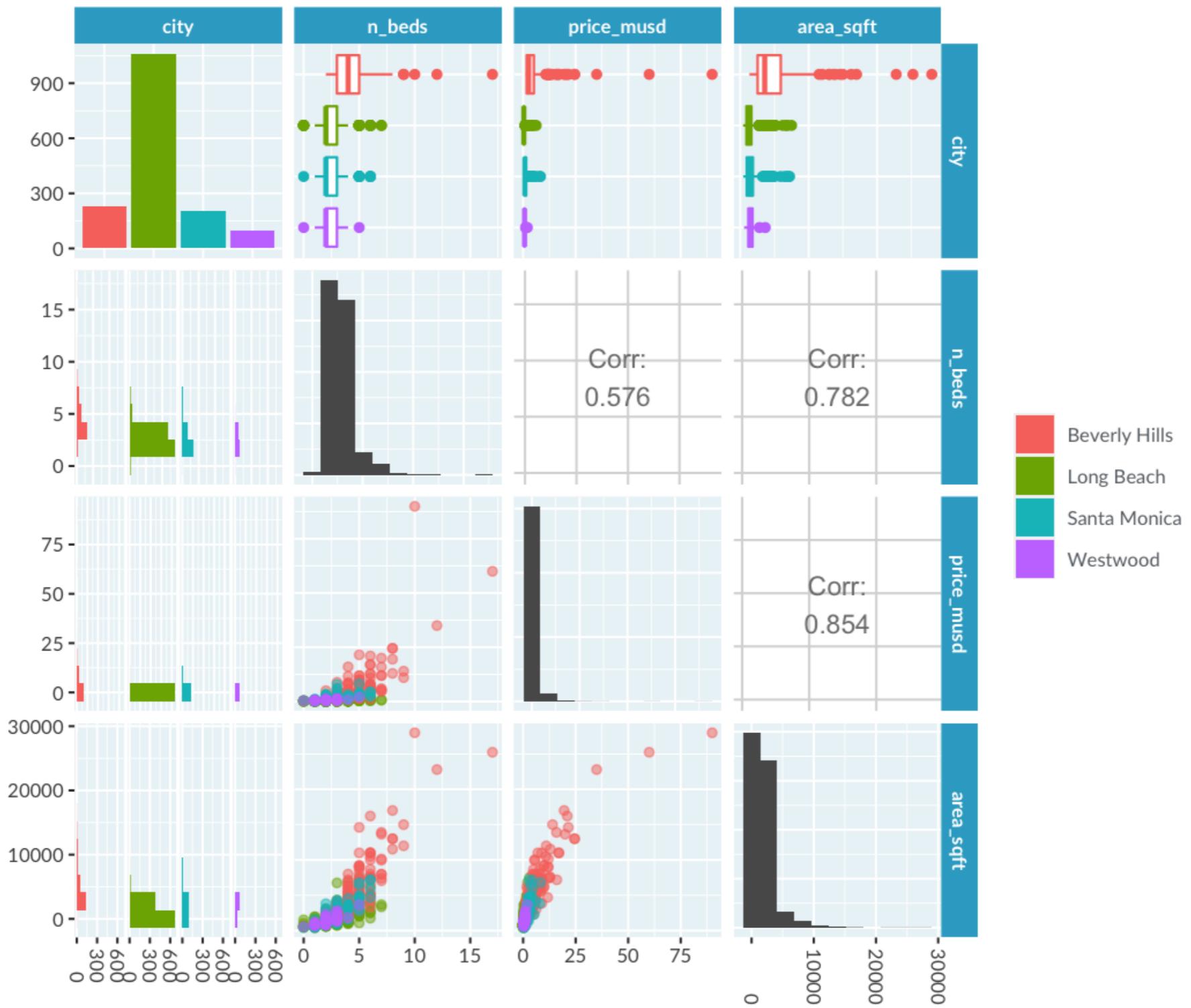
UNDERSTANDING DATA VISUALIZATION



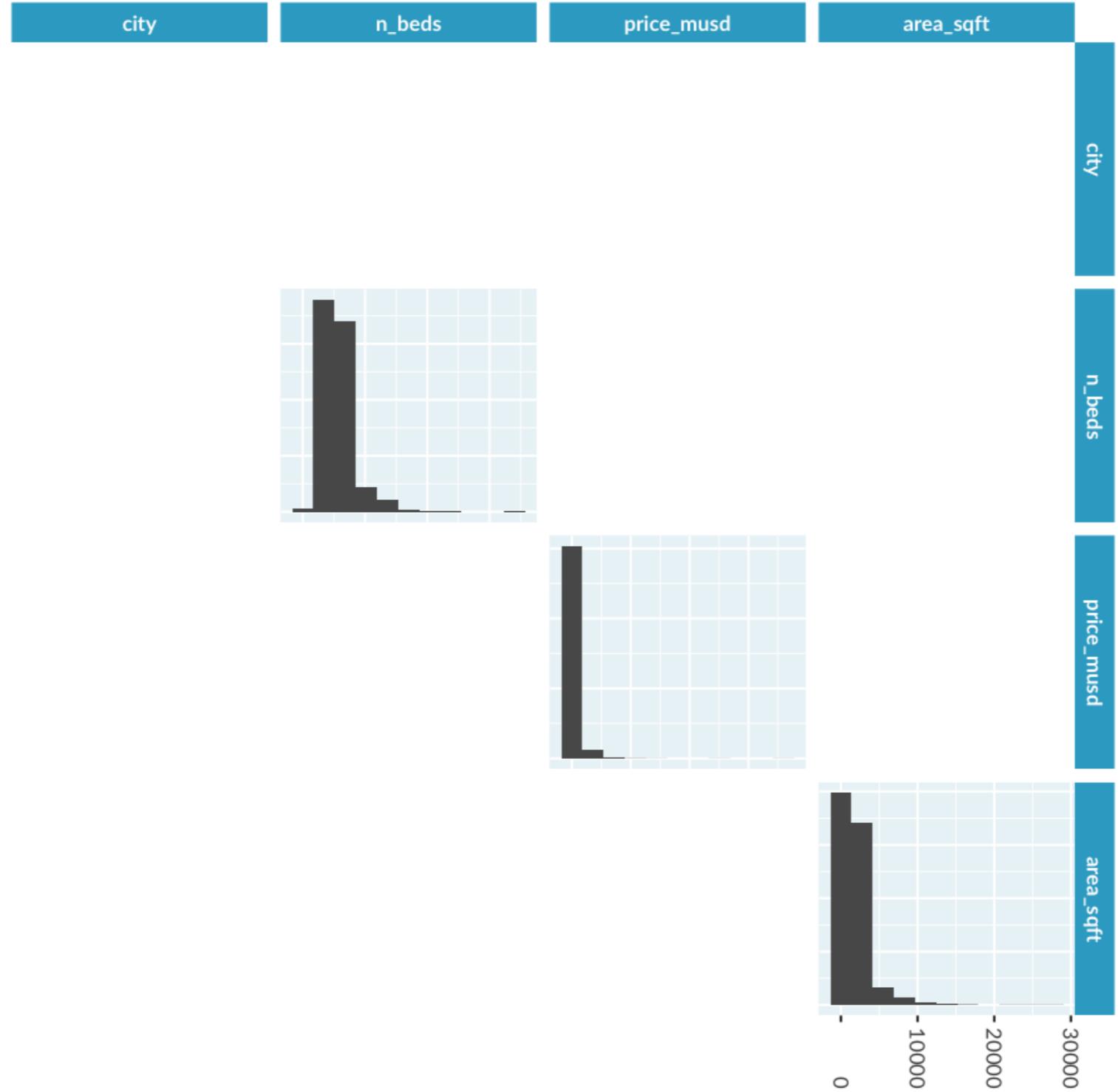
Richie Cotton
Data Evangelist

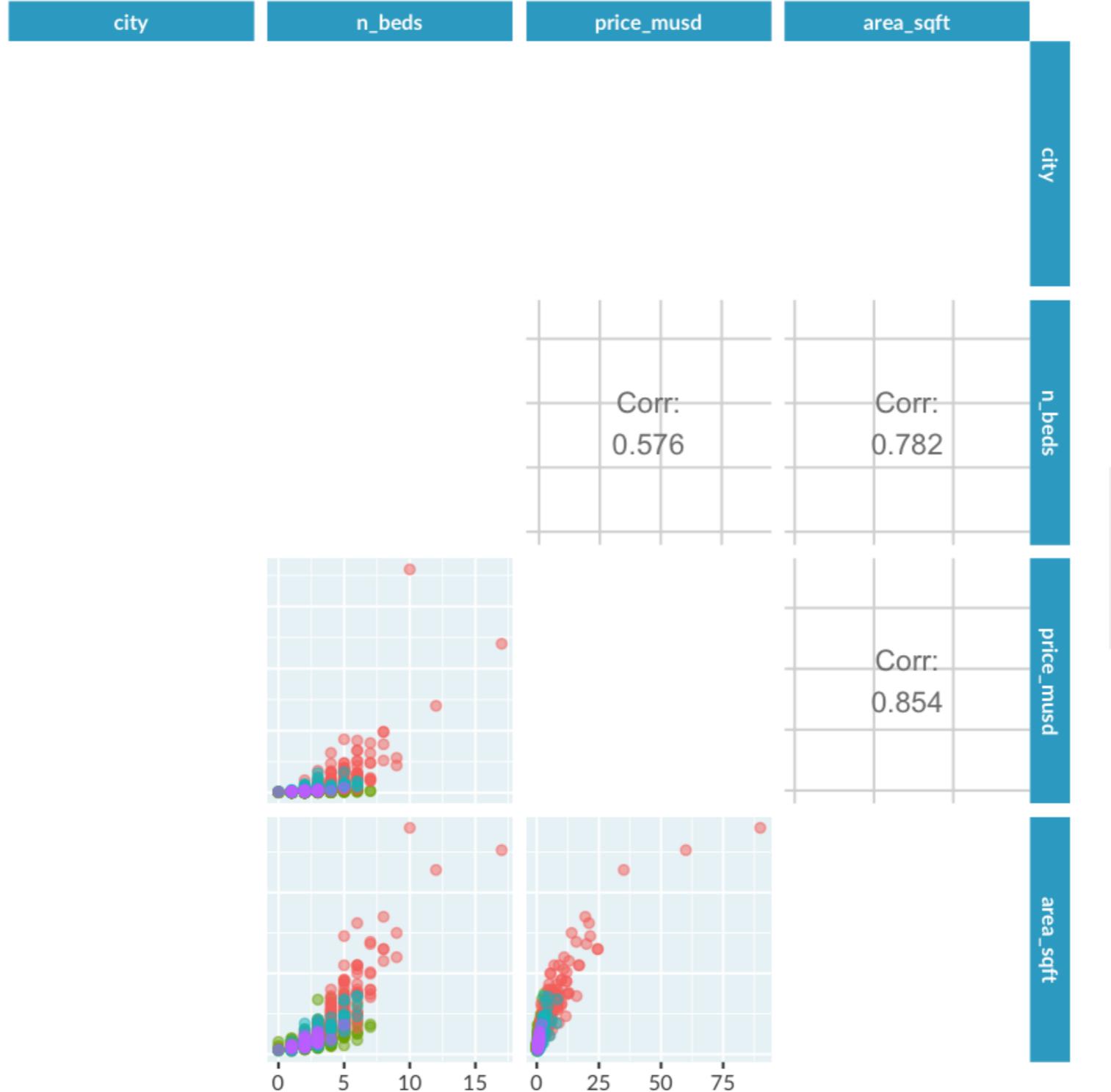
When should you use a pair plot?

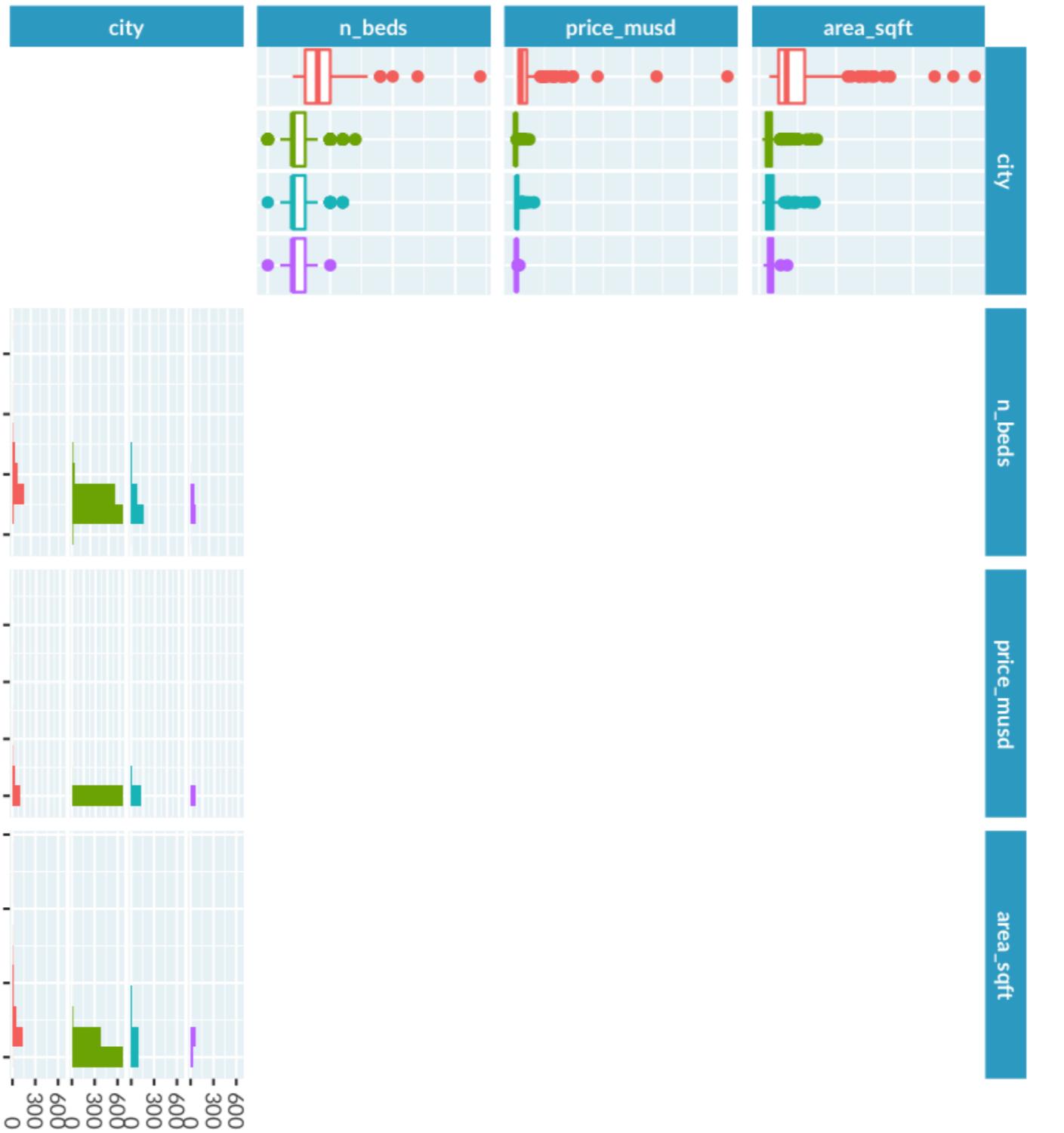
- You have up to ten variables (either continuous, categorical, or a mix).
- You want to see the distribution for each variable.
- You want to see the relationship between each pair of variables.

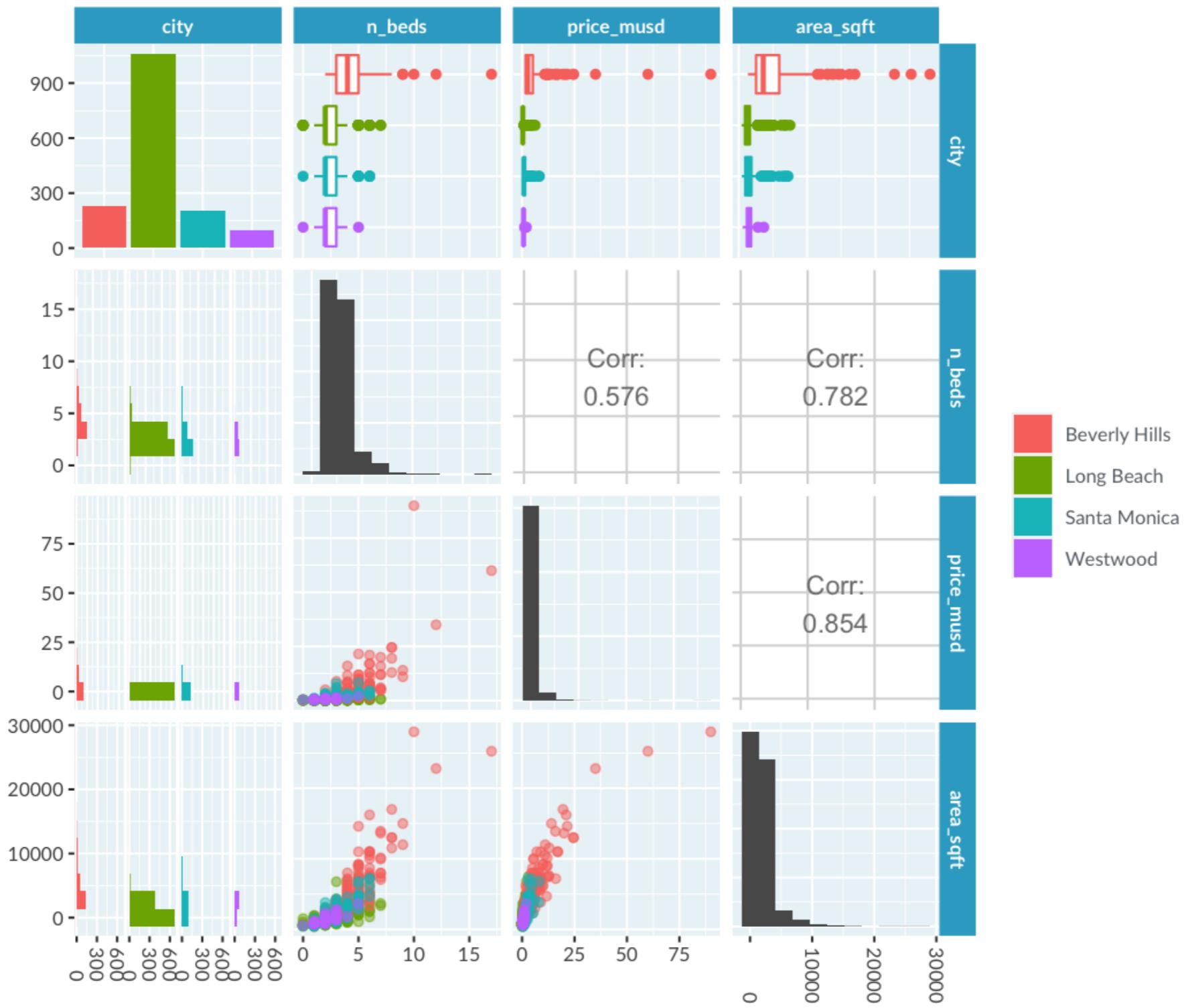


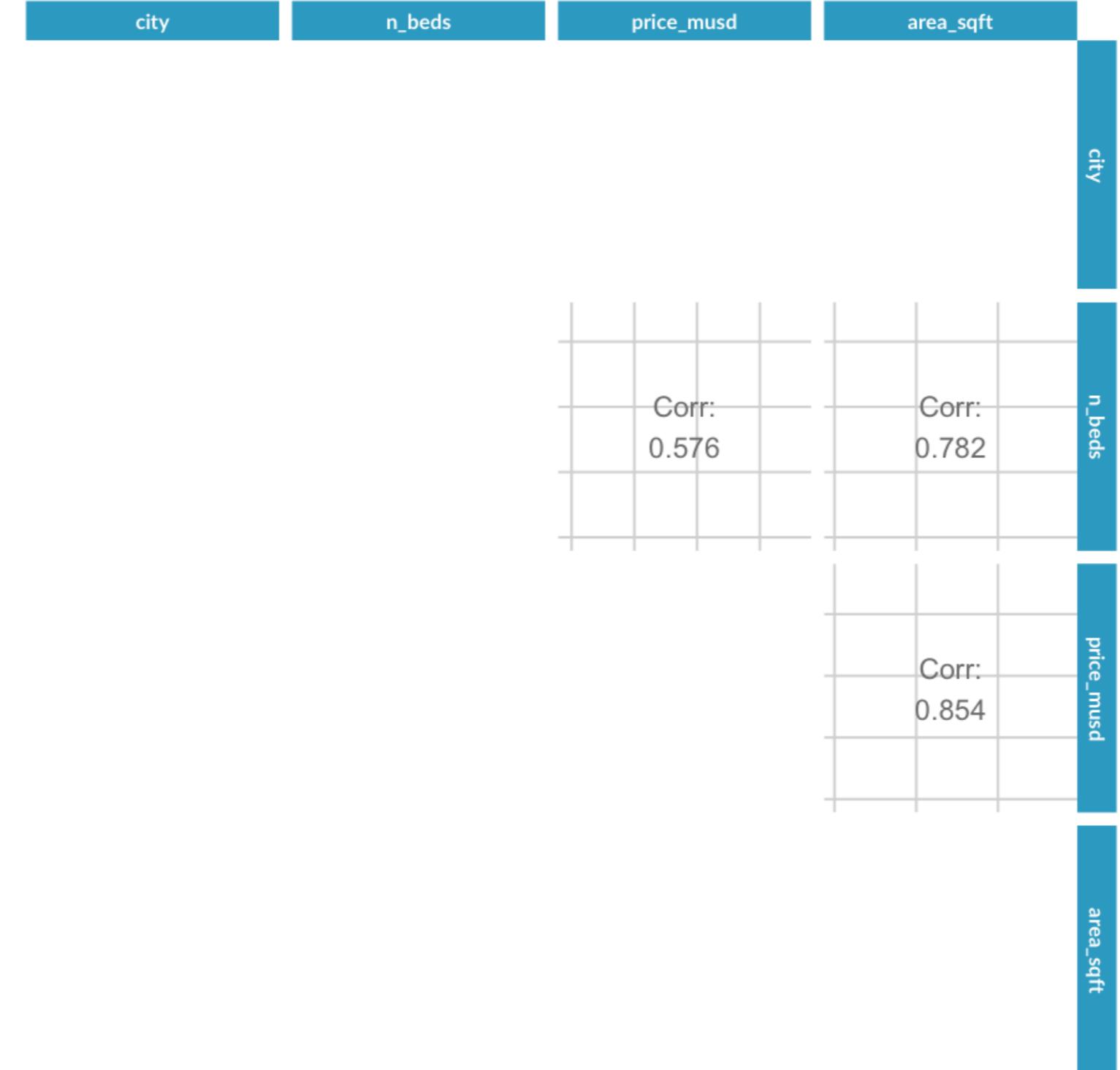






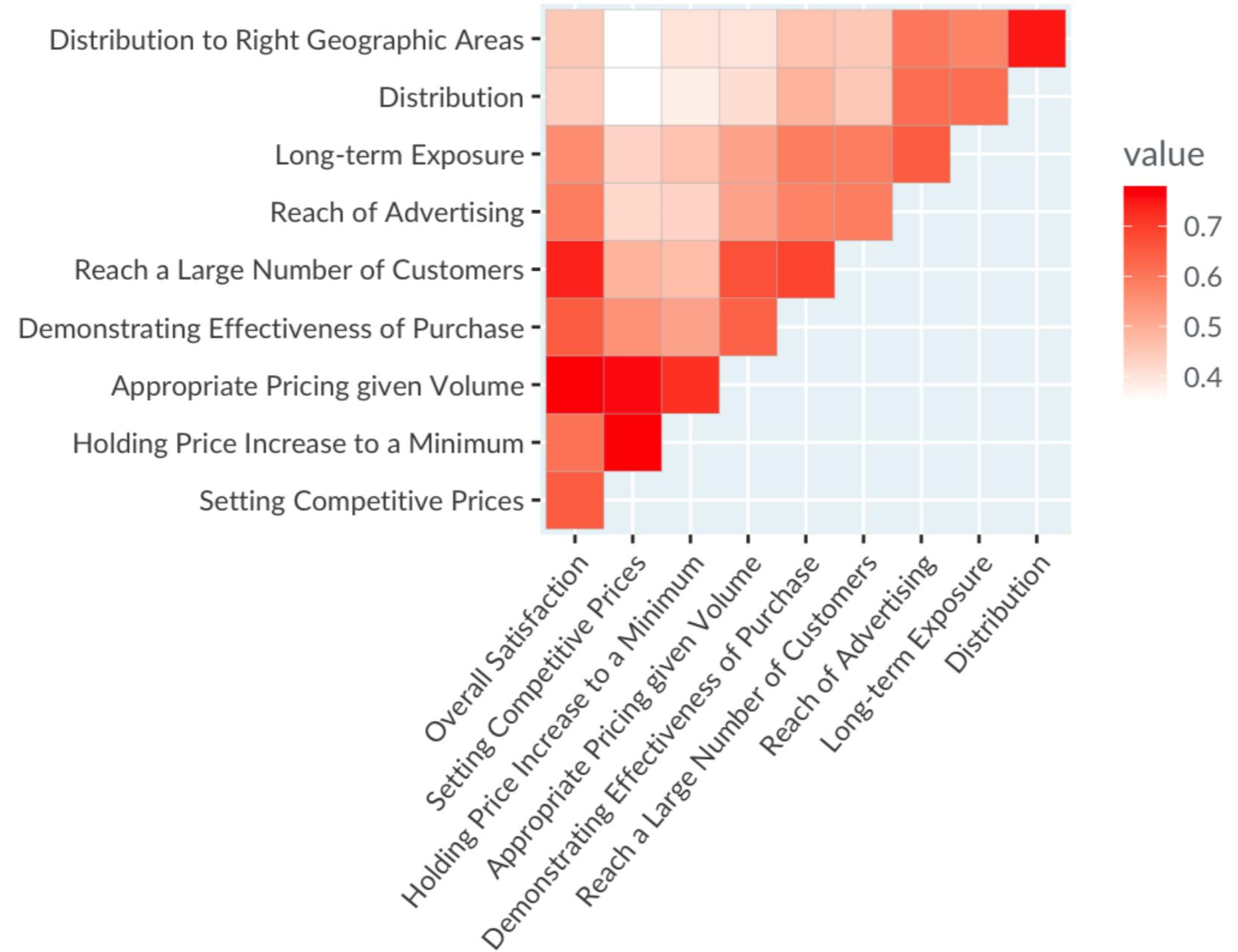






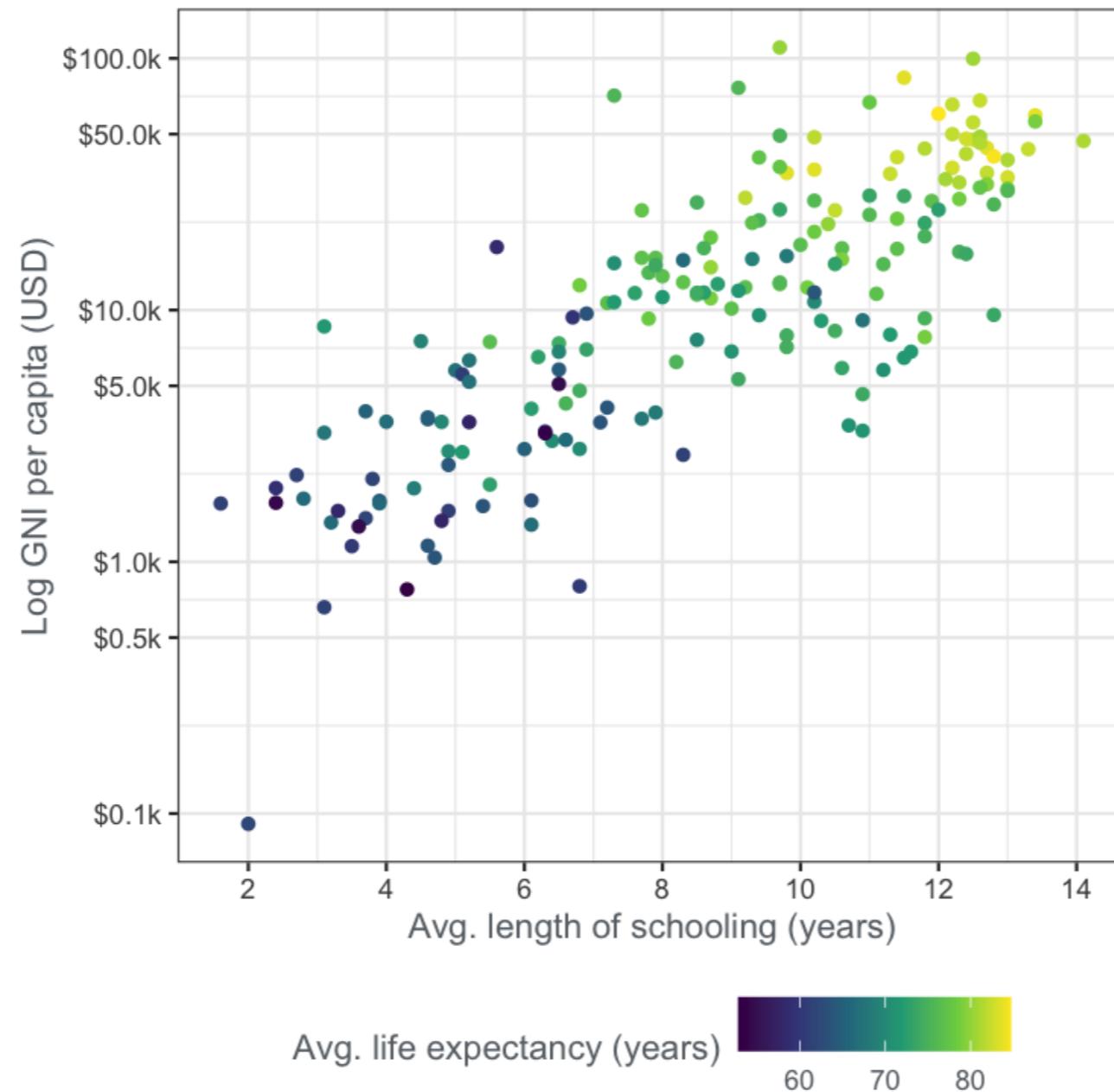
When should you use a correlation heatmap?

- You have lots of continuous variables.
- You want to a simple overview of how each pair of variables is related.



¹ Rossi, Allenby, and McCulloch (2005). Bayesian Statistics & Marketing

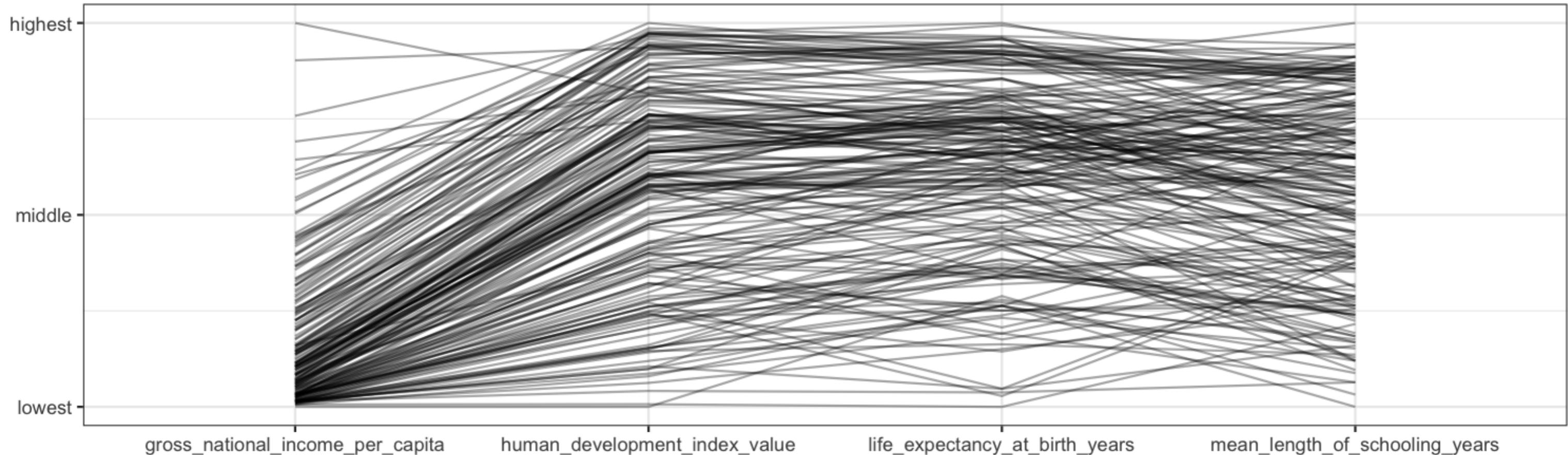
The United Nations dataset again

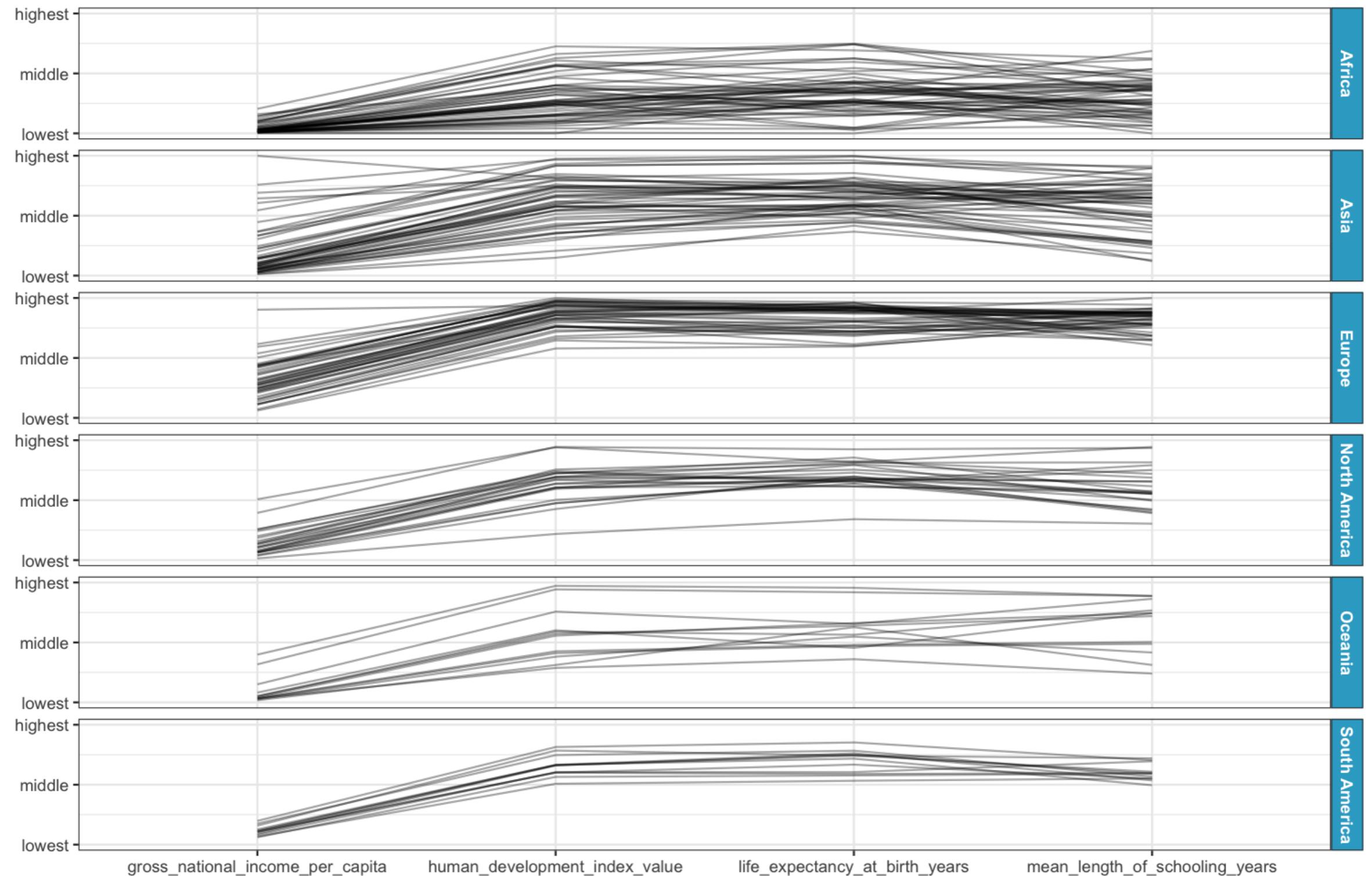


When should you use a parallel coordinates plot?

- You have lots of continuous variables.
- You want to find patterns across these variables, or
- You want to visualize clusters of observations.

A parallel coordinates plot





Let's practice!

UNDERSTANDING DATA VISUALIZATION