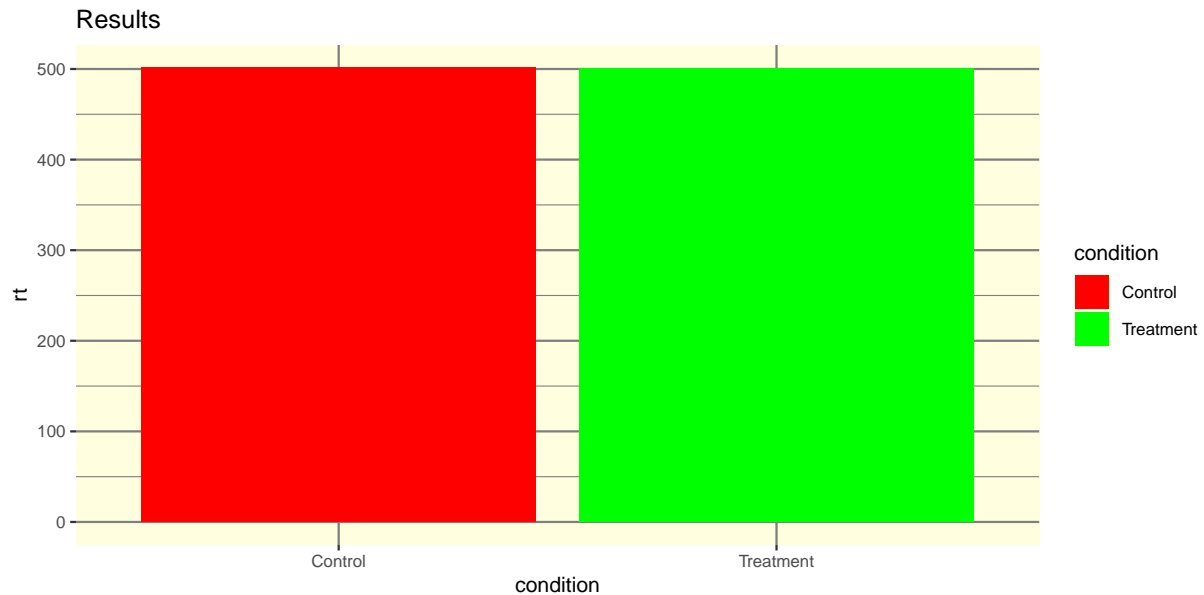# Perception & Design

PSY 410: Data Science for Psychology

Dr. Sara Weston

2026-04-22

How we see data

What's wrong with this?

. . .

Red-green palette (colorblind-unfriendly). No informative title. Tiny text. Gridline noise. Bars hiding the data.

Today we learn **why your brain rejects bad figures** — and how to design ones that work.

Your brain processes some visual features before you even think

Some visual features are processed almost instantly — before conscious attention kicks in:

Attribute	Example
Color	A red dot among blue dots
Size	A large circle among small ones
Position	A point far from the others
Shape	A triangle among circles
Orientation	A tilted bar among vertical bars

These are your tools for drawing the viewer's eye.

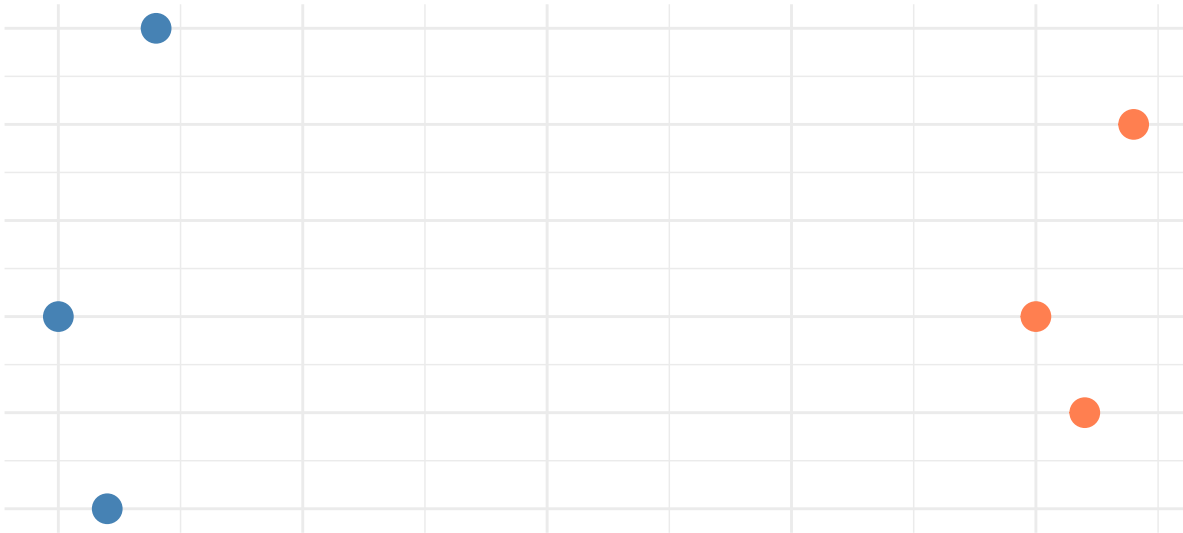**Preattentive in action**
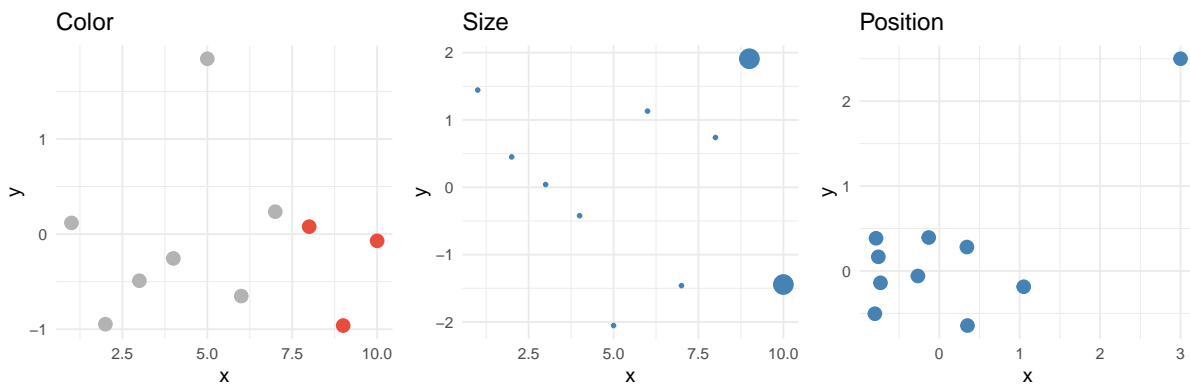
Your eye goes to the red points, the big points, and the outlier — instantly.

Gestalt principles (brief)

The brain groups things automatically:

- **Proximity** — things close together feel like a group
- **Similarity** — things that look alike feel like a group
- **Enclosure** — things inside a border feel like a group

Proximity: two clusters feel like two groups

Color theory

Three types of color palettes

Type	When to use	Example
Sequential	One continuous variable (low $\rightarrow$ high)	Blues, viridis
Diverging	Values relative to a midpoint	Red–white–blue
Qualitative	Categorical groups (no order)	Set1, tab10

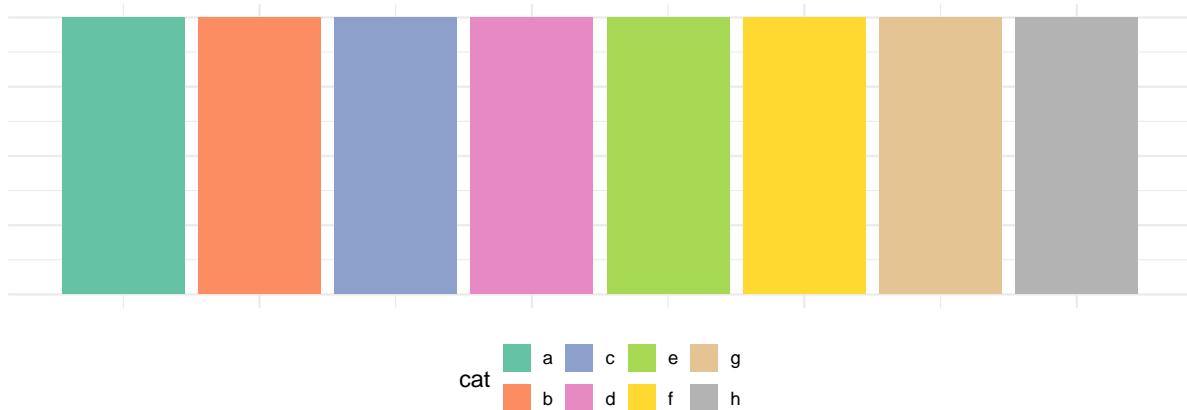Using the wrong type is one of the most common visualization mistakes.

Sequential palettes

viridis (sequential) – great for continuous data
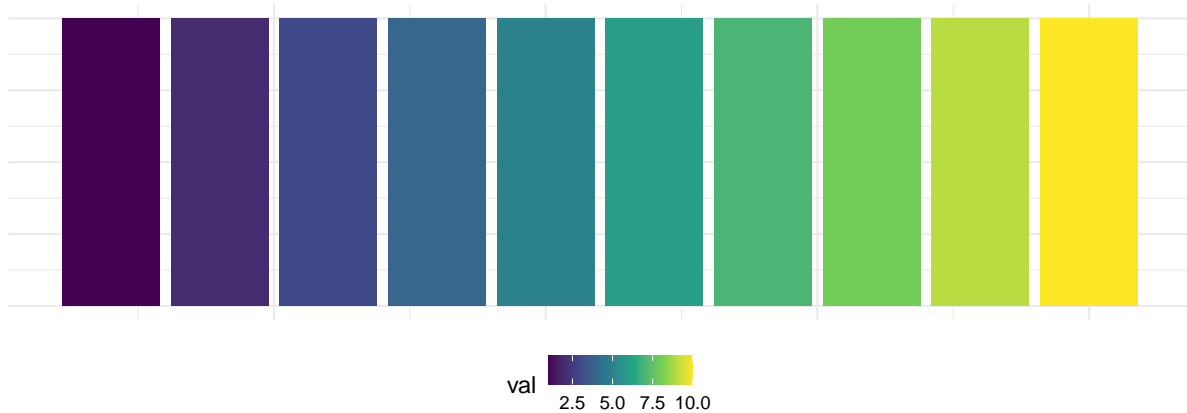
Qualitative palettes

Set2 (qualitative) – for categorical groups


Colorblind-friendly choices

About 8% of men have some form of color vision deficiency. Red-green is the most common.

```
# viridis is colorblind-safe AND sequential
scale_fill_viridis_d()   # discrete
scale_fill_viridis_c()   # continuous

# ColorBrewer palettes designed for colorblindness
scale_fill_brewer(palette = "Set2")
```

```
# Or set colors manually with safe choices
scale_fill_manual(values = c("#0072B2", "#E69F00", "#009E73"))
# (blue, orange, green - distinguishable for most color vision types)
```
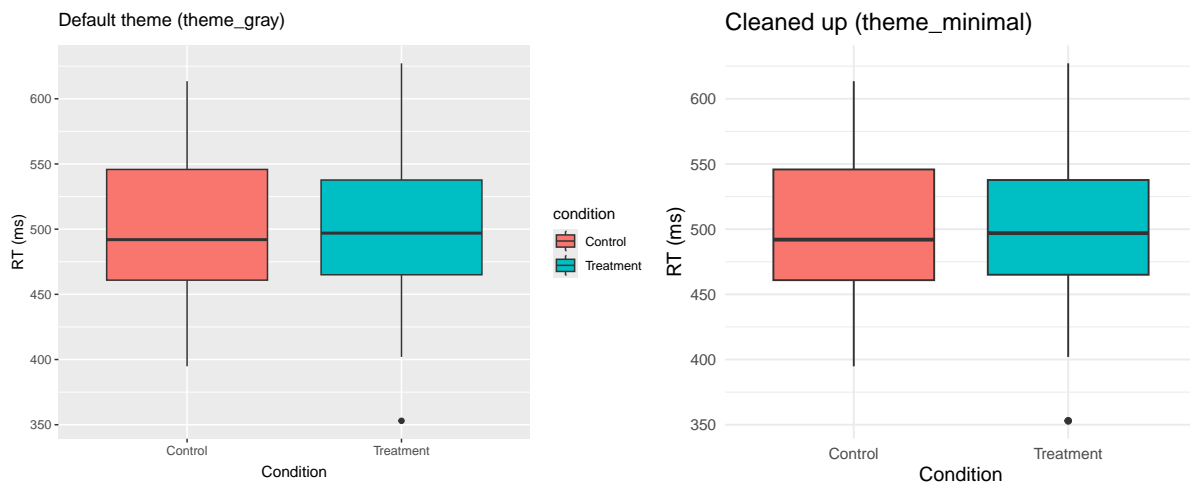
A bad color choice vs a good one
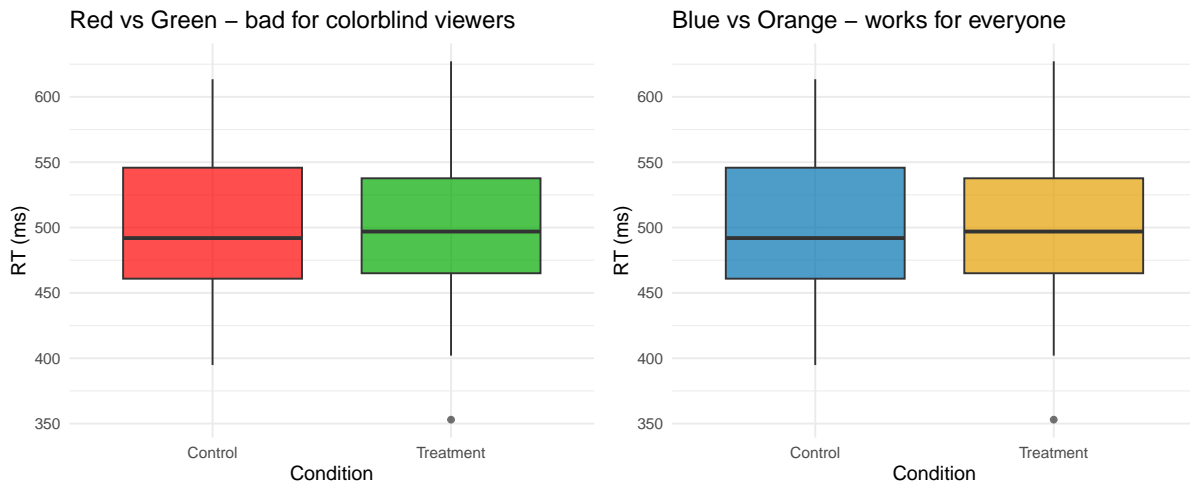
Decluttering

The default is cluttered

ggplot2's default theme adds a lot of visual noise. Compare:

The declutter checklist

Ask: does this element help the reader understand the data?

Remove

Element	Why
Gray background	Noise, no info
Gridlines (most)	Distraction
Redundant legend	X-axis says it
Generic axis labels	Add units instead

Keep

Element	Why
Title + subtitle	Orients the reader
Caption	Source, N, error bar type
Meaningful color	Highlights comparisons

theme() for fine control

```
theme(
  legend.position = "none",                # Remove legend
  axis.ticks = element_blank(),            # Remove tick marks
  panel.grid = element_blank(),            # Remove gridlines
  plot.title = element_text(face = "bold"),   # Bold the title
  axis.text = element_text(size = 10),     # Font size
  plot.caption = element_text(hjust = 0)   # Left-align caption
)
```

Pair coding break

Your turn: 10 minutes

I'll put a cluttered graph on screen. With a partner, **rewrite it** to follow the design principles we just covered.

Remove at least 5 unnecessary elements. Make it tell a clear story.

Your turn: 10 minutes

```r
# The cluttered version - fix this!
reaction_data |>
  ggplot(aes(x = condition, y = rt, fill = condition, color = condition, size = accuracy)) +
  geom_point() +
  geom_boxplot(alpha = 0.3) +
  scale_fill_manual(values = c("Control" = "red", "Treatment" = "green")) +
  scale_color_manual(values = c("Control" = "red", "Treatment" = "green")) +
  labs(x = "condition", y = "rt") +
  ggtitle("data") +
  theme_gray()
```
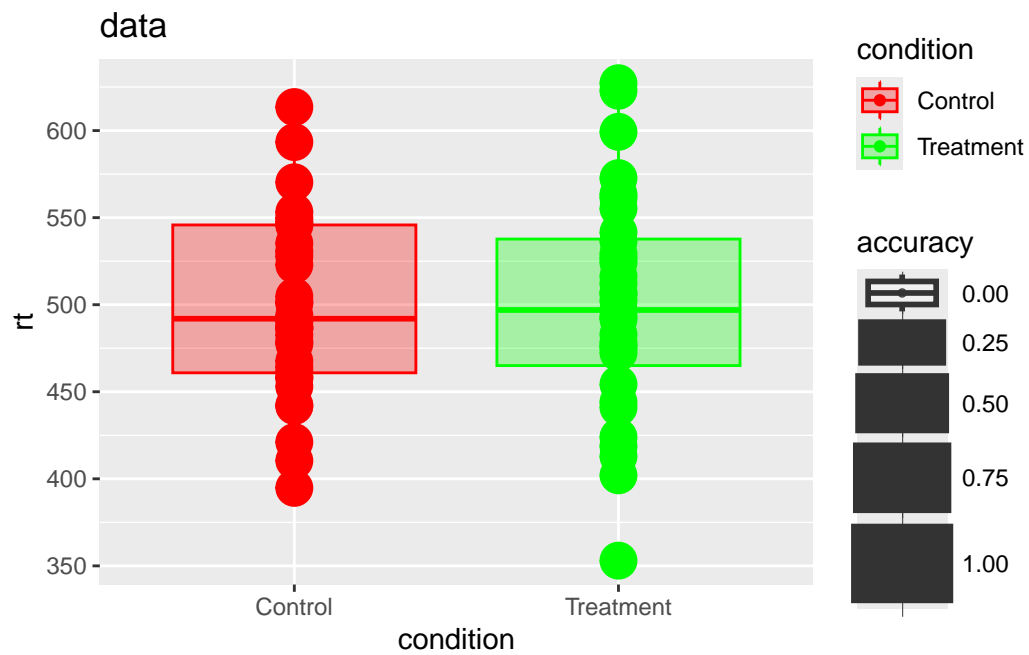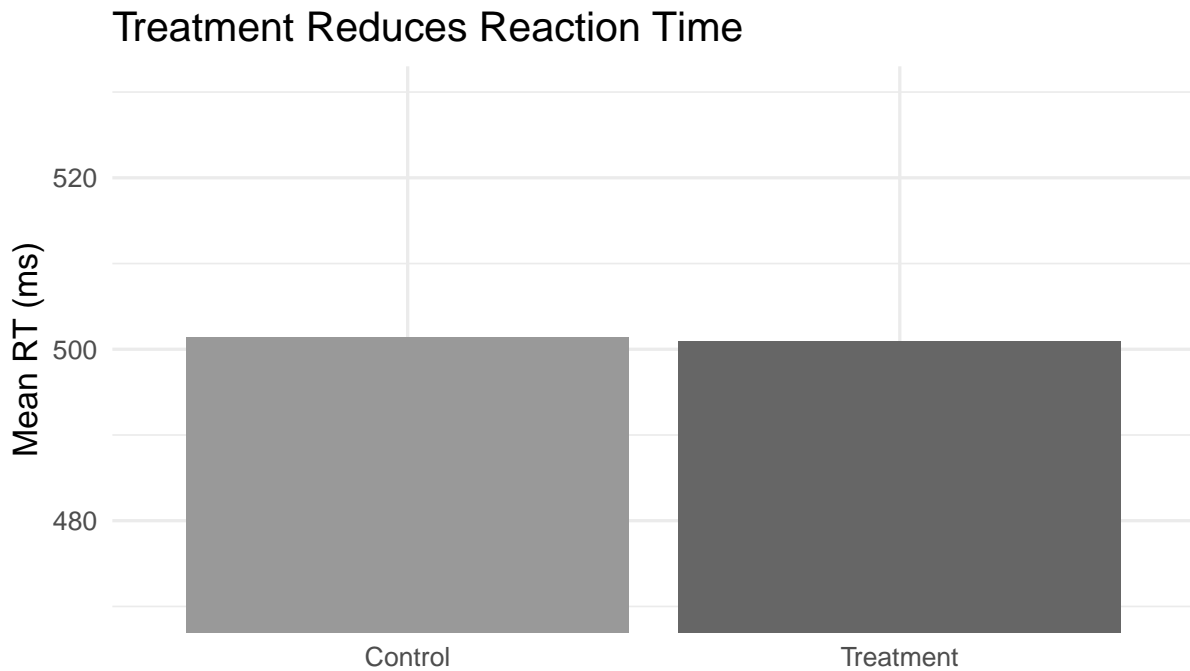
Before we move on

Upload your code to Canvas for participation credit. Paste what you have into today's in-class submission — it doesn't need to work perfectly.
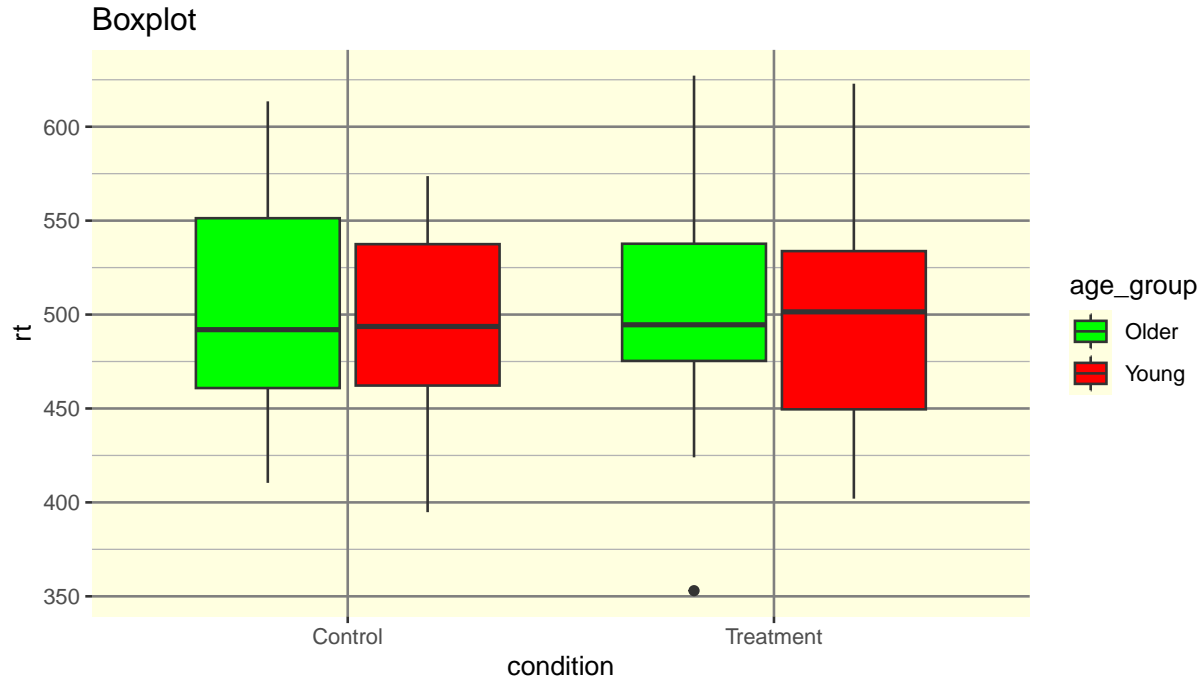
Critiquing bad graphs

What's wrong here? (1 of 3)

Treatment Reduces Reaction Time

. . .

The problem: The y-axis starts at 470, not 0. The difference looks massive — but it's only ~40 ms. A truncated axis exaggerates the effect.
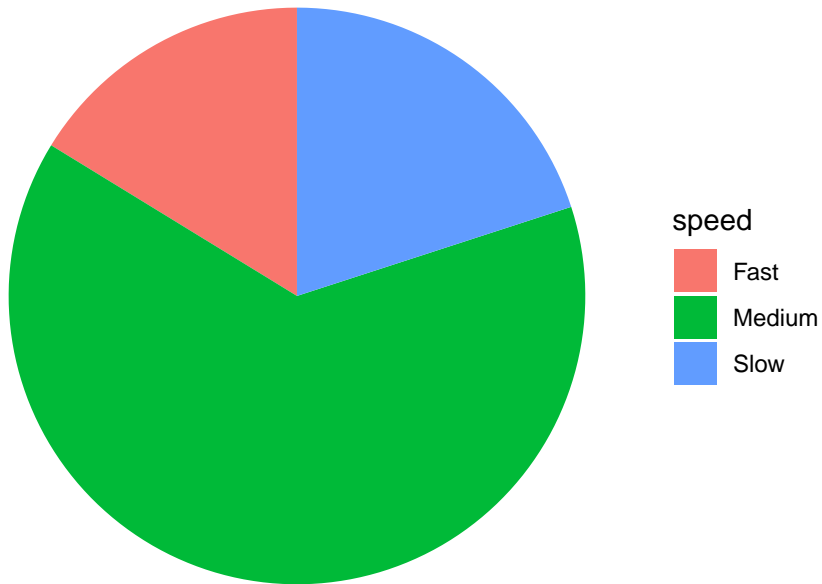
What's wrong here? (2 of 3)

Boxplot

. . .

The problems: Red-green palette (colorblind-unfriendly). Title says "Boxplot" (a label, not a finding). Variable names as labels. Distracting background color and gridlines.

Speed Categories

. . .

The problems: Continuous data was binned into arbitrary categories, then displayed as a pie chart. We lost the actual reaction times, the condition comparison, and the ability to see distributions. A histogram or density plot would show far more.

Get a head start

Assignment 4 preview

Assignment 4 will ask you to:

1. Create a **"bad" version** of a figure — deliberately violate design principles
2. Create a **"good" version** following what we covered today
3. Create a **colorblind-accessible version**

Assignment 4 preview

Start experimenting now:

- Take the `reaction_data` dataset
- Make the worst possible version of a figure
- Then make it great
- What did you change?

Wrapping up

Design principles checklist

☐ Is the geom appropriate for the data type?
☐ Are colors colorblind-friendly?
☐ Is the legend necessary? (Remove if redundant)
☐ Are axis labels informative (not just variable names)?
☐ Is there a title that states the main point?
☐ Is there visual clutter to remove?
☐ Do error bars have a caption explaining what they show?

Before next class

Read:

- [R4DS Ch 10: Exploratory data analysis](#) (sections 10.1–10.4)

Practice:

- Try the "bad vs good" exercise on your own
- Find a graph online and identify ways to improve it

Key takeaways

1. **Preattentive attributes** guide the eye — use them intentionally
2. **Match your palette type** to your data (sequential, diverging, qualitative)
3. **Design for colorblindness** — always
4. **Declutter ruthlessly** — every element must earn its place
5. **Critical evaluation** — always ask whether each element adds information

The one thing to remember

You're not designing for yourself. You're designing for a reader who will look at your figure for five seconds.

Next time: Exploratory Data Analysis