# Publication-quality figures

*Eleutherodactylus* Research Course · Fall 2026

Tom Devitt

2026-10-28

# Today

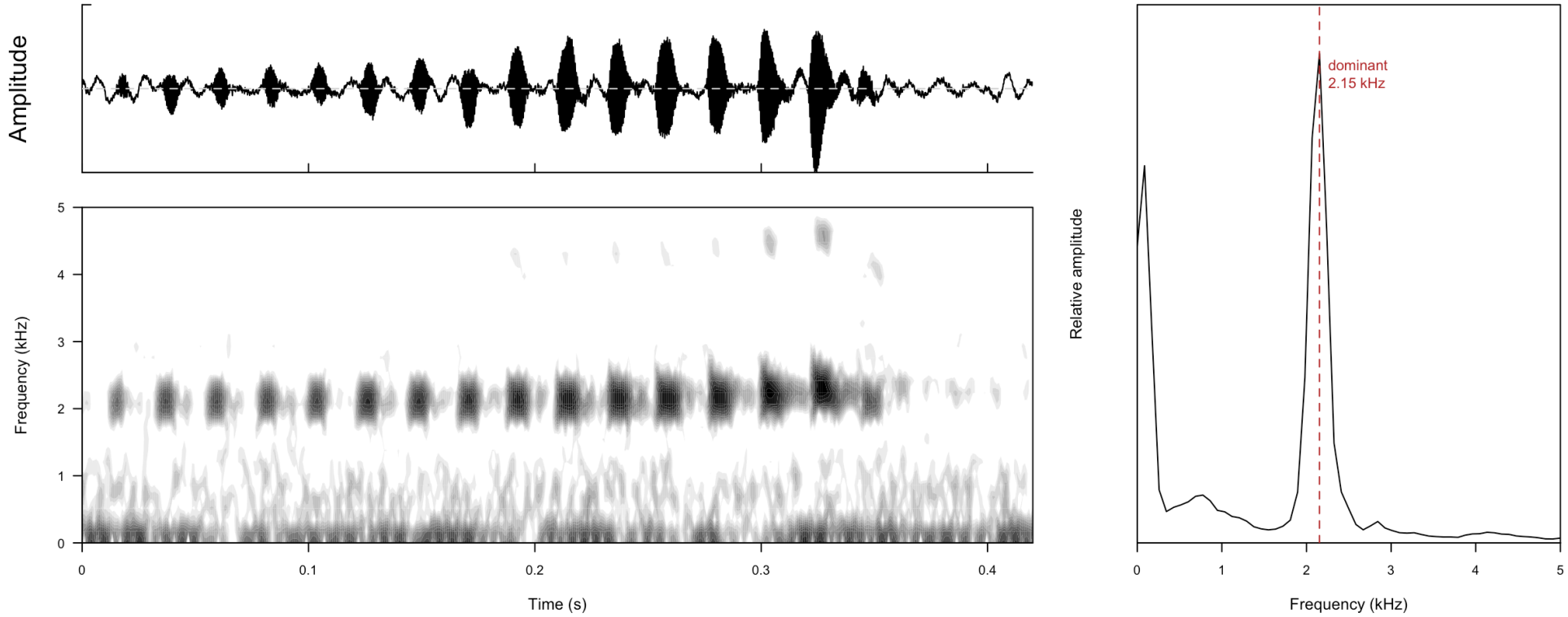
About 15 minutes of lesson, then a figure sprint (a short, focused work session where each team builds and improves one figure).

A figure is **publication-quality** when a reader who has not seen your notebook can understand it without help.

Source: Claus Wilke, *Fundamentals of Data Visualization* (free online).

If your data is not clean yet, keep working on the notebook. Milestone 2 is still due Mon Nov 2.

# A good figure: one call, three views

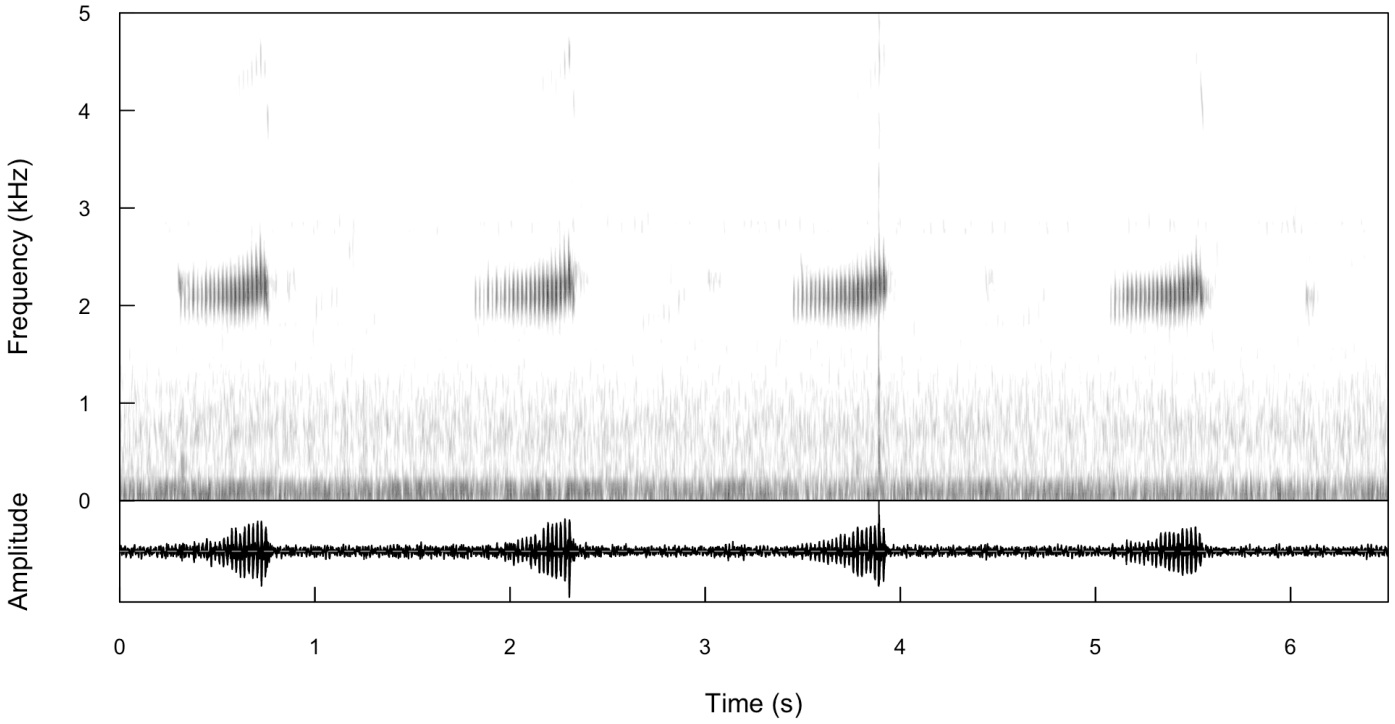



*E. marnockii*, Brackenridge Field Lab, one call, 0.42 s.

# What makes it work

- **Every axis has words and units:** “Frequency (kHz)”, “Time (s)”
- **The three panels share an axis:** the oscillogram and spectrogram line up in time, so you can match a pulse to its dark blob
- **The measurement is marked on the figure:** the red line and label show the dominant frequency, 2.15 kHz
- **Grayscale only:** no color is needed, so nothing is lost in print
- **No clutter:** no title banner, no grid, no legend box

# A good figure: a series of calls




*E. marnockii*, four calls in 6.5 s.

The time axis is long enough to show the **rhythm** (the gap between calls), and the oscillogram strip keeps each call tied to its spectrogram.

# 1. Start with one question

Before you plot, finish this sentence: “This figure shows that _____.”

| Question | Plot that fits |
|---|---|
| How are the values spread out? | histogram or density plot |
| How do groups compare? | boxplot, or points with the mean marked |
| Do two measurements go together? | scatterplot |

Show the individual calls when you can. With a few calls per species, a reader wants to see every one.

# 2. Label so the figure stands alone

- Axis titles in words with units: “Dominant frequency (kHz)”, not `dom_freq`
- Species names in italics, spelled the way the paper will spell them
- Say how many calls (n) each group has
- Text large enough to read when the figure is half a page wide

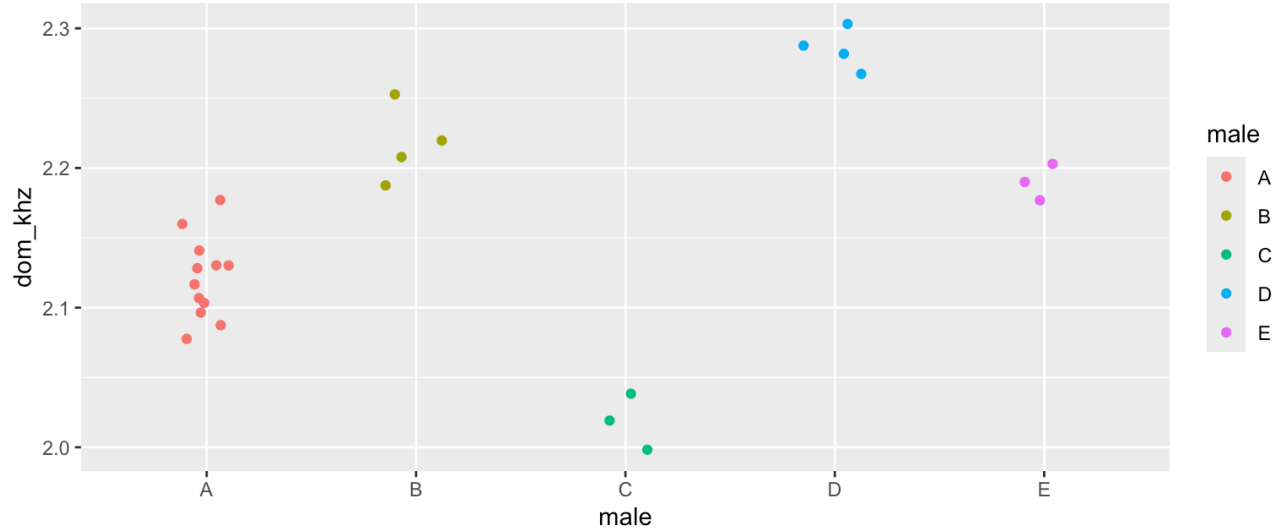
```
labs(x = "Dominant frequency (kHz)", y = "Call duration (s)")
```

# 3. Color, clutter, and order

- Use color only to carry information (such as species), and pick a palette that stays distinct in grayscale and for color-blind readers (`scale_color_viridis_d()`)
- Remove what carries no information: heavy gridlines, grey backgrounds, boxes around legends (`theme_minimal()` or `theme_classic()` is a good start)
- Put groups in a meaningful order (by mean value, or in the order of the tree), not alphabetical

# Before: the default plot

```
1 ggplot(calls, aes(x = male, y = dom_khz, color = male)) +
2   geom_jitter(width = 0.15)
```

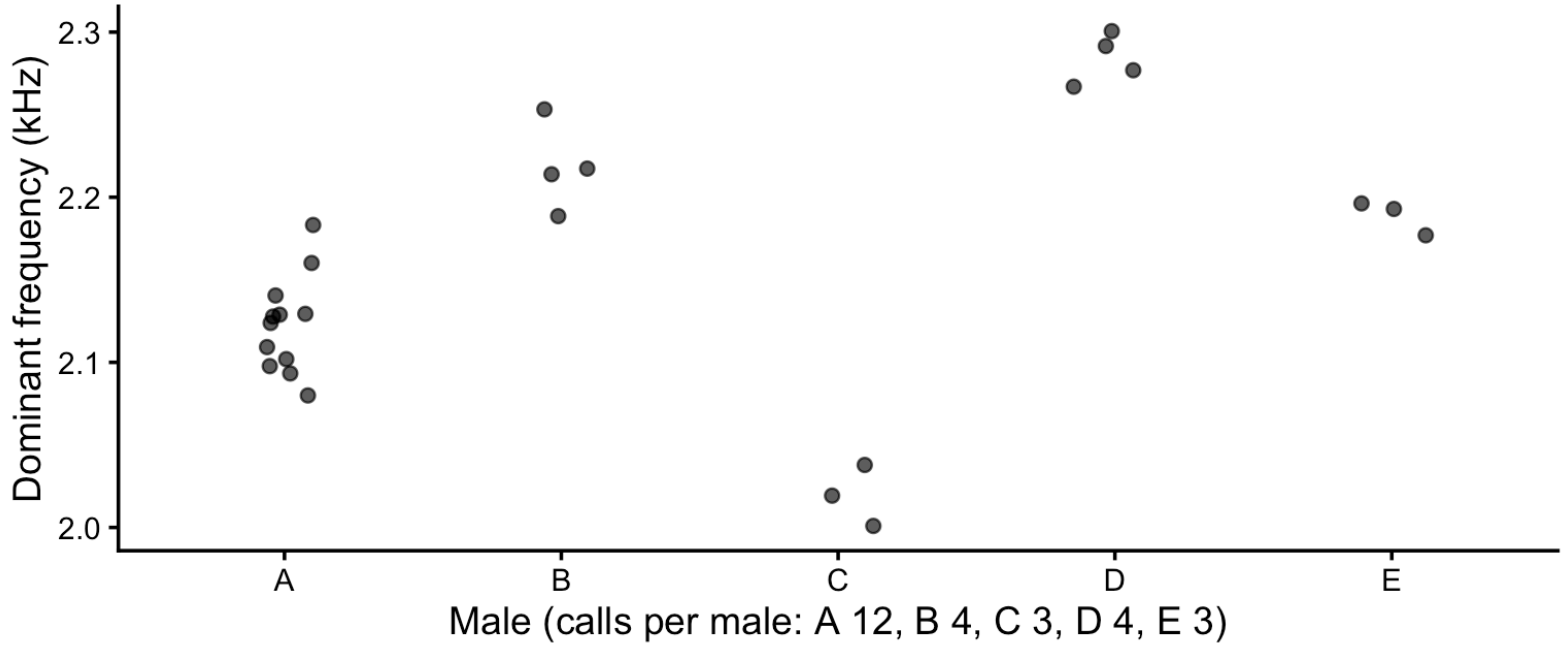



*Simulated marnockii-like data, five males.* What would a reader ask?

- What is `dom_khz`? Is it a frequency, and in what units?
- How many calls and how many males?
- Why is every male a different color when the x axis already names them?

# After: the same data, fixed

```
ggplot(calls, aes(x = male, y = dom_khz)) +
  geom_jitter(width = 0.15, alpha = 0.7, size = 2) +
  labs(x = "Male (calls per male: A 12, B 4, C 3, D 4, E 3)",
       y = "Dominant frequency (kHz)") +
  theme_classic(base_size = 14)
```

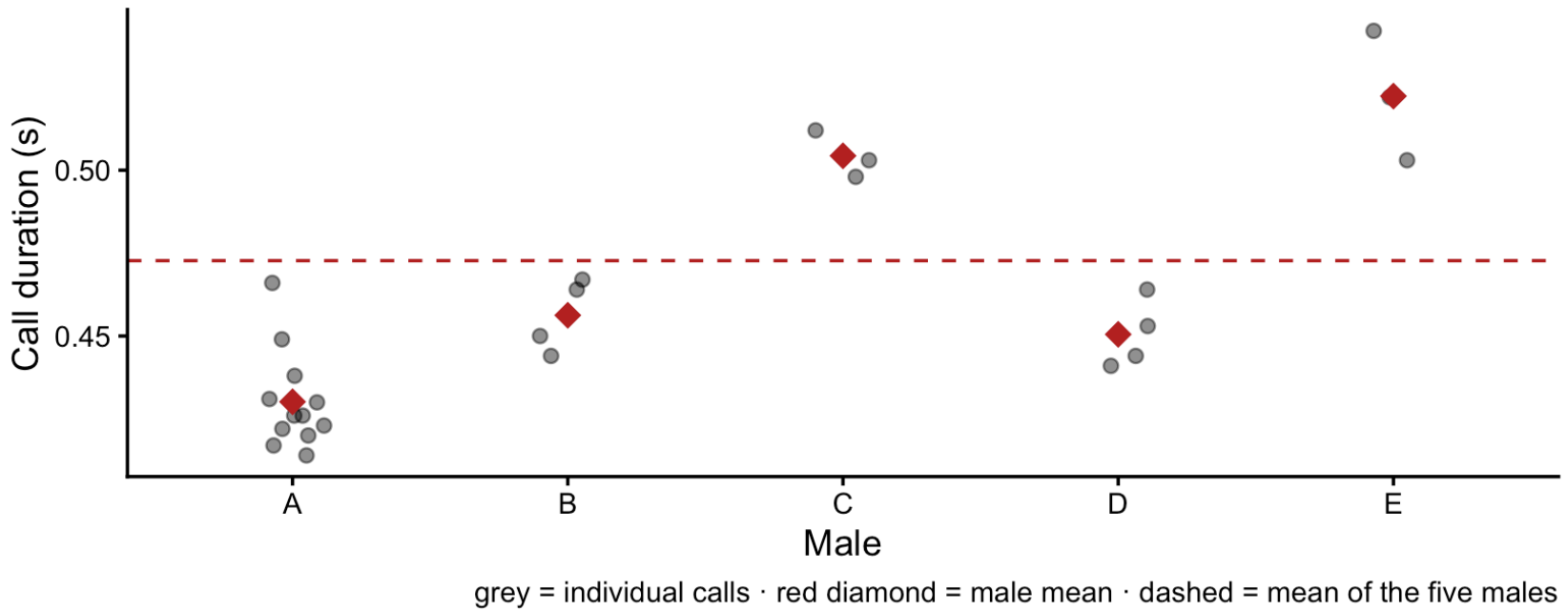



Same points. Words and units on the axes, n stated, color removed because it carried no information, gridlines and grey panel gone.

# A caption that explains every mark

```
by_male <- calls |> summarise(duration_s = mean(duration_s), .by = male)
mean_of_males <- mean(by_male$duration_s)

ggplot(calls, aes(male, duration_s)) +
  geom_jitter(width = 0.12, height = 0, alpha = 0.5, size = 2.2) +
  geom_point(data = by_male, color = "firebrick", size = 4.5, shape = 18) +
  geom_hline(yintercept = mean_of_males, linetype = 2, color = "firebrick") +
  labs(x = "Male", y = "Call duration (s)",
       caption = "grey = individual calls · red diamond = male mean · dashed = mean of the five males") +
  theme_classic(base_size = 14)
```




ut-fri-biodiversity.github.io/eleuth-calls

# Why that figure works

*Simulated marnockii-like data, five males.*

- Every call is drawn, so the reader sees the data and not just a summary
- Calls from one male are not independent, so the male means (not the 26 calls) are what you describe the species with
- The caption says what each symbol is, so the figure needs no one to explain it

# 4. Save it at the size it will be printed

A screenshot is not a figure file. Save from code, so the figure can be remade whenever the data changes:

```
ggsave("figs/fig1-call-duration.png",
       width = 6, height = 4, units = "in", dpi = 300)
```

Width 6 in is about one page column. Set the size once, then check that the text is still readable.

# Sprint

Each team picks **one** figure from your notebook and works it through this checklist:

1. One sentence: “This figure shows that _____.”
2. Axis titles have words and units
3. n is shown or stated
4. Color and order mean something
5. Saved with `ggsave()` at a set size

Mentors and Tom will visit each team. In the last 5 minutes, post your best figure and its sentence from step 1 on Canvas.

# What comes next

| Date | What |
|---|---|
| Mon Nov 2 | **Milestone 2** due: clean data and QC notebook |
| Mon Nov 16 | **Milestone 3** due: complete figures and results (Nov 9 and 11 are a catch-up week) |

# Credits

*Prepared with Claude Code (Claude Sonnet 5.5, Anthropic): first draft of the deck from the course schedule and Wilke's chapter topics; spectrogram examples are from the Sep 9 sound deck; the dot plots use simulated data.*